



47

INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(51) International Patent Classification 7 : C12N 15/56, 15/63, 1/21, 9/24, 15/11		A1	(11) International Publication Number: WO 00/52178 (43) International Publication Date: 8 September 2000 (08.09.00)
(21) International Application Number: PCT/US00/03542		(81) Designated States: AE, AL, AM, AT, AU, AZ, BA, BB, BG, BR, BY, CA, CH, CN, CR, CU, CZ, DE, DK, DM, EE, ES, FI, GB, GD, GE, GH, GM, HR, HU, ID, IL, IN, IS, JP, KE, KG, KP, KR, KZ, LC, LK, LR, LS, LT, LU, LV, MA, MD, MG, MK, MN, MW, MX, NO, NZ, PL, PT, RO, RU, SD, SE, SG, SI, SK, SL, TJ, TM, TR, TT, TZ, UA, UG, US, UZ, VN, YU, ZA, ZW, ARIPO patent (GH, GM, KE, LS, MW, SD, SL, SZ, TZ, UG, ZW), Eurasian patent (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), European patent (AT, BE, CH, CY, DE, DK, ES, FI, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE), OAPI patent (BF, BJ, CF, CG, CI, CM, GA, GN, GW, ML, MR, NE, SN, TD, TG).	
(22) International Filing Date: 14 February 2000 (14.02.00)			
(30) Priority Data: 09/258,892 1 March 1999 (01.03.99) US			
(71) Applicants (for all designated States except US): INSIGHT STRATEGY & MARKETING LTD. [IL/IL]; P.O. Box 2128, 76121 Rehovot (IL). HADASIT MEDICAL RESEARCH SERVICES & DEVELOPMENT LTD. [IL/IL]; P.O. Box 12000, 91120 Jerusalem (IL).			
(71) Applicant (for TJ only): FRIEDMAN, Mark, M. [US/IL]; 1 Alharizi Street, 43406 Raanana (IL).			
(72) Inventors; and		Published	
(75) Inventors/Applicants (for US only): PECKER, Iris [IL/IL]; 42 Wolfson Street, 75203 Rishon Lezion (IL). VLODAVSKY, Israel [IL/IL]; 34 Arbel Street, 90805 Mevaseret Zion (IL). FEINSTEIN, Elena [IL/IL]; 12/29 Hahagana Street, 76214 Rehovot (IL).		With international search report. Before the expiration of the time limit for amending the claims and to be republished in the event of the receipt of amendments.	
(74) Common Representative: FRIEDMAN, Mark, M.; c/o Castorina, Anthony, Suite 207, 2001 Jefferson Davis Highway, Arlington, VA 22202 (US).			
(54) Title: POLYNUCLEOTIDE ENCODING A POLYPEPTIDE HAVING HEPARANASE ACTIVITY AND EXPRESSION OF SAME IN GENETICALLY MODIFIED CELLS			
(57) Abstract			
<p>A polynucleotide (<i>hpa</i>) encoding a polypeptide having heparanase activity, vectors including same, genetically modified cells expressing heparanase, a recombinant protein having heparanase activity and antisense oligonucleotides and constructs for modulating heparanase expression.</p>			

FOR THE PURPOSES OF INFORMATION ONLY

Codes used to identify States party to the PCT on the front pages of pamphlets publishing international applications under the PCT.

AL	Albania	ES	Spain	LS	Lesotho	SI	Slovenia
AM	Armenia	FI	Finland	LT	Lithuania	SK	Slovakia
AT	Austria	FR	France	LU	Luxembourg	SN	Senegal
AU	Australia	GA	Gabon	LV	Latvia	SZ	Swaziland
AZ	Azerbaijan	GB	United Kingdom	MC	Monaco	TD	Chad
BA	Bosnia and Herzegovina	GE	Georgia	MD	Republic of Moldova	TG	Togo
BB	Barbados	GH	Ghana	MG	Madagascar	TJ	Tajikistan
BE	Belgium	GN	Guinea	MK	The former Yugoslav Republic of Macedonia	TM	Turkmenistan
BF	Burkina Faso	GR	Greece	ML	Mali	TR	Turkey
BG	Bulgaria	HU	Hungary	MN	Mongolia	TT	Trinidad and Tobago
BJ	Benin	IE	Ireland	MR	Mauritania	UA	Ukraine
BR	Brazil	IL	Israel	MW	Malawi	UG	Uganda
BY	Belarus	IS	Iceland	MX	Mexico	US	United States of America
CA	Canada	IT	Italy	NE	Niger	UZ	Uzbekistan
CF	Central African Republic	JP	Japan	NL	Netherlands	VN	Viet Nam
CG	Congo	KE	Kenya	NO	Norway	YU	Yugoslavia
CH	Switzerland	KG	Kyrgyzstan	NZ	New Zealand	ZW	Zimbabwe
CI	Côte d'Ivoire	KP	Democratic People's Republic of Korea	PL	Poland		
CM	Cameroon	KR	Republic of Korea	PT	Portugal		
CN	China	KZ	Kazakhstan	RO	Romania		
CU	Cuba	LC	Saint Lucia	RU	Russian Federation		
CZ	Czech Republic	LI	Liechtenstein	SD	Sudan		
DE	Germany	LK	Sri Lanka	SE	Sweden		
DK	Denmark	LR	Liberia	SG	Singapore		

**POLYNUCLEOTIDE ENCODING A POLYPEPTIDE HAVING
HEPARANASE ACTIVITY AND EXPRESSION OF SAME IN
GENETICALLY MODIFIED CELLS**

5 FIELD AND BACKGROUND OF THE INVENTION

The present invention relates to a polynucleotide, referred to hereinbelow as *hpa*, encoding a polypeptide having heparanase activity, vectors (nucleic acid constructs) including same and genetically modified cells expressing heparanase. The invention further relates to a recombinant 10 protein having heparanase activity and to antisense oligonucleotides, constructs and ribozymes for down regulating heparanase activity. In addition, the invention relates to heparanase promoter sequences and their uses.

Heparan sulfate proteoglycans: Heparan sulfate proteoglycans (HSPG) are ubiquitous macromolecules associated with the cell surface and extra cellular matrix (ECM) of a wide range of cells of vertebrate and invertebrate tissues (1-4). The basic HSPG structure includes a protein core to which several linear heparan sulfate chains are covalently attached. These polysaccharide chains are typically composed of repeating hexuronic 15 and D-glucosamine disaccharide units that are substituted to a varying extent with N- and O-linked sulfate moieties and N-linked acetyl groups (1-4). Studies on the involvement of ECM molecules in cell attachment, growth and differentiation revealed a central role of HSPG in embryonic morphogenesis, angiogenesis, neurite outgrowth and tissue repair (1-5). 20 HSPG are prominent components of blood vessels (3). In large blood vessels they are concentrated mostly in the intima and inner media, whereas in capillaries they are found mainly in the subendothelial basement membrane where they support proliferating and migrating endothelial cells and stabilize the structure of the capillary wall. The ability of HSPG to interact with ECM macromolecules such as collagen, laminin and 25 fibronectin, and with different attachment sites on plasma membranes suggests a key role for this proteoglycan in the self-assembly and insolubility of ECM components, as well as in cell adhesion and locomotion. Cleavage of the heparan sulfate (HS) chains may therefore 30 result in degradation of the subendothelial ECM and hence may play a decisive role in extravasation of blood-borne cells. HS catabolism is observed in inflammation, wound repair, diabetes, and cancer metastasis, suggesting that enzymes which degrade HS play important roles in 35 pathologic processes. Heparanase activity has been described in activated

immune system cells and highly metastatic cancer cells (6-8), but research has been handicapped by the lack of biologic tools to explore potential causative roles of heparanase in disease conditions.

Involvement of Heparanase in Tumor Cell Invasion and

Metastasis: Circulating tumor cells arrested in the capillary beds of different organs must invade the endothelial cell lining and degrade its underlying basement membrane (BM) in order to invade into the extravascular tissue(s) where they establish metastasis (9, 10). Metastatic tumor cells often attach at or near the intercellular junctions between adjacent endothelial cells. Such attachment of the metastatic cells is followed by rupture of the junctions, retraction of the endothelial cell borders and migration through the breach in the endothelium toward the exposed underlying BM (9). Once located between endothelial cells and the BM, the invading cells must degrade the subendothelial glycoproteins and proteoglycans of the BM in order to migrate out of the vascular compartment. Several cellular enzymes (e.g., collagenase IV, plasminogen activator, cathepsin B, elastase, etc.) are thought to be involved in degradation of BM (10). Among these enzymes is an endo- β -D-glucuronidase (heparanase) that cleaves HS at specific intrachain sites (6, 8, 11). Expression of a HS degrading heparanase was found to correlate with the metastatic potential of mouse lymphoma (11), fibrosarcoma and melanoma (8) cells. Moreover, elevated levels of heparanase were detected in sera from metastatic tumor bearing animals and melanoma patients (8) and in tumor biopsies of cancer patients (12).

The control of cell proliferation and tumor progression by the local microenvironment, focusing on the interaction of cells with the extracellular matrix (ECM) produced by cultured corneal and vascular endothelial cells, was investigated previously by the present inventors. This cultured ECM closely resembles the subendothelium *in vivo* in its morphological appearance and molecular composition. It contains collagens (mostly type III and IV, with smaller amounts of types I and V), proteoglycans (mostly heparan sulfate- and dermatan sulfate- proteoglycans, with smaller amounts of chondroitin sulfate proteoglycans), laminin, fibronectin, entactin and elastin (13, 14). The ability of cells to degrade HS in the cultured ECM was studied by allowing cells to interact with a metabolically sulfate labeled ECM, followed by gel filtration (Sephadex 6B) analysis of degradation products released into the culture medium (11). While intact HSPG are eluted next to the void volume of the column

(Kav<0.2, Mr ~ 0.5x10⁶), labeled degradation fragments of HS side chains are eluted more toward the V_t of the column (0.5<kav<0.8, Mr =5-7x10³) (11).

The heparanase inhibitory effect of various non-anticoagulant species of heparin that might be of potential use in preventing extravasation of blood-borne cells was also investigated by the present inventors. Inhibition of heparanase was best achieved by heparin species containing 16 sugar units or more and having sulfate groups at both the N and O positions. While O-desulfation abolished the heparanase inhibiting effect of heparin, O-sulfated, N-acetylated heparin retained a high inhibitory activity, provided that the N-substituted molecules had a molecular size of about 4,000 daltons or more (7). Treatment of experimental animals with heparanase inhibitors (e.g., non-anticoagulant species of heparin) markedly reduced (>90%) the incidence of lung metastases induced by B16 melanoma, Lewis lung carcinoma and mammary adenocarcinoma cells (7, 8, 16). Heparin fractions with high and low affinity to anti-thrombin III exhibited a comparable high anti-metastatic activity, indicating that the heparanase inhibiting activity of heparin, rather than its anticoagulant activity, plays a role in the anti-metastatic properties of the polysaccharide (7).

Heparanase activity in the urine of cancer patients: In an attempt to further elucidate the involvement of heparanase in tumor progression and its relevance to human cancer, urine samples for heparanase activity were screened (16a). Heparanase activity was detected in the urine of some, but not all, cancer patients. High levels of heparanase activity were determined in the urine of patients with an aggressive metastatic disease and there was no detectable activity in the urine of healthy donors.

Heparanase activity was also found in the urine of 20% of normal and microalbuminuric insulin dependent diabetes mellitus (IDDM) patients, most likely due to diabetic nephropathy, the most important single disorder leading to renal failure in adults.

Possible involvement of heparanase in tumor angiogenesis: Fibroblast growth factors are a family of structurally related polypeptides characterized by high affinity to heparin (17). They are highly mitogenic for vascular endothelial cells and are among the most potent inducers of neovascularization (17, 18). Basic fibroblast growth factor (bFGF) has been extracted from the subendothelial ECM produced *in vitro* (19) and from basement membranes of the cornea (20), suggesting that ECM may serve as

a reservoir for bFGF. Immunohistochemical staining revealed the localization of bFGF in basement membranes of diverse tissues and blood vessels (21). Despite the ubiquitous presence of bFGF in normal tissues, endothelial cell proliferation in these tissues is usually very low, suggesting that bFGF is somehow sequestered from its site of action. Studies on the interaction of bFGF with ECM revealed that bFGF binds to HSPG in the ECM and can be released in an active form by HS degrading enzymes (15, 20, 22). It was demonstrated that heparanase activity expressed by platelets, mast cells, neutrophils, and lymphoma cells is involved in release of active bFGF from ECM and basement membranes (23), suggesting that heparanase activity may not only function in cell migration and invasion, but may also elicit an indirect neovascular response. These results suggest that the ECM HSPG provides a natural storage depot for bFGF and possibly other heparin-binding growth promoting factors (24, 25). Displacement of bFGF from its storage within basement membranes and ECM may therefore provide a novel mechanism for induction of neovascularization in normal and pathological situations.

Recent studies indicate that heparin and HS are involved in binding of bFGF to high affinity cell surface receptors and in bFGF cell signaling (26, 27). Moreover, the size of HS required for optimal effect was similar to that of HS fragments released by heparanase (28). Similar results were obtained with vascular endothelial cells growth factor (VEGF) (29), suggesting the operation of a dual receptor mechanism involving HS in cell interaction with heparin-binding growth factors. It is therefore proposed that restriction of endothelial cell growth factors in ECM prevents their systemic action on the vascular endothelium, thus maintaining a very low rate of endothelial cells turnover and vessel growth. On the other hand, release of bFGF from storage in ECM as a complex with HS fragment, may elicit localized endothelial cell proliferation and neovascularization in processes such as wound healing, inflammation and tumor development (24, 25).

Expression of heparanase by cells of the immune system: Heparanase activity correlates with the ability of activated cells of the immune system to leave the circulation and elicit both inflammatory and autoimmune responses. Interaction of platelets, granulocytes, T and B lymphocytes, macrophages and mast cells with the subendothelial ECM is associated with degradation of HS by a specific heparanase activity (6). The enzyme is released from intracellular compartments (e.g., lysosomes,

specific granules, etc.) in response to various activation signals (e.g., thrombin, calcium ionophore, immune complexes, antigens, mitogens, etc.), suggesting its regulated involvement in inflammation and cellular immunity.

5 *Some of the observations regarding the heparanase enzyme were reviewed in reference No. 6 and are listed hereinbelow:*

First, a proteolytic activity (plasminogen activator) and heparanase participate synergistically in sequential degradation of the ECM HSPG by inflammatory leukocytes and malignant cells.

10 Second, a large proportion of the platelet heparanase exists in a latent form, probably as a complex with chondroitin sulfate. The latent enzyme is activated by tumor cell-derived factor(s) and may then facilitate cell invasion through the vascular endothelium in the process of tumor metastasis.

15 Third, release of the platelet heparanase from α -granules is induced by a strong stimulant (i.e., thrombin), but not in response to platelet activation on ECM.

20 Fourth, the neutrophil heparanase is preferentially and readily released in response to a threshold activation and upon incubation of the cells on ECM.

25 Fifth, contact of neutrophils with ECM inhibited release of noxious enzymes (proteases, lysozyme) and oxygen radicals, but not of enzymes (heparanase, gelatinase) which may enable diapedesis. This protective role of the subendothelial ECM was observed when the cells were stimulated with soluble factors but not with phagocytosable stimulants.

Sixth, intracellular heparanase is secreted within minutes after exposure of T cell lines to specific antigens.

30 Seventh, mitogens (Con A, LPS) induce synthesis and secretion of heparanase by normal T and B lymphocytes maintained *in vitro*. T lymphocyte heparanase is also induced by immunization with antigen *in vivo*.

Eighth, heparanase activity is expressed by pre-B lymphomas and B-lymphomas, but not by plasmacytomas and resting normal B lymphocytes.

35 Ninth, heparanase activity is expressed by activated macrophages during incubation with ECM, but there was little or no release of the enzyme into the incubation medium. Similar results were obtained with human myeloid leukemia cells induced to differentiate to mature macrophages.

Tenth, T-cell mediated delayed type hypersensitivity and experimental autoimmunity are suppressed by low doses of heparanase inhibiting non-anticoagulant species of heparin (30).

Eleventh, heparanase activity expressed by platelets, neutrophils and metastatic tumor cells releases active bFGF from ECM and basement membranes. Release of bFGF from storage in ECM may elicit a localized neovascular response in processes such as wound healing, inflammation and tumor development.

Twelfth, among the breakdown products of the ECM generated by heparanase is a tri-sulfated disaccharide that can inhibit T-cell mediated inflammation *in vivo* (31). This inhibition was associated with an inhibitory effect of the disaccharide on the production of biologically active TNF α by activated T cells *in vitro* (31).

Other potential therapeutic applications: Apart from its involvement in tumor cell metastasis, inflammation and autoimmunity, mammalian heparanase may be applied to modulate: bioavailability of heparin-binding growth factors (15); cellular responses to heparin-binding growth factors (e.g., bFGF, VEGF) and cytokines (IL-8) (31a, 29); cell interaction with plasma lipoproteins (32); cellular susceptibility to certain viral and some bacterial and protozoa infections (33, 33a, 33b); and disintegration of amyloid plaques (34). Heparanase may thus prove useful for conditions such as wound healing, angiogenesis, restenosis, atherosclerosis, inflammation, neurodegenerative diseases and viral infections. Mammalian heparanase can be used to neutralize plasma heparin, as a potential replacement of protamine. Anti-heparanase antibodies may be applied for immunodetection and diagnosis of micrometastases, autoimmune lesions and renal failure in biopsy specimens, plasma samples, and body fluids. Common use in basic research is expected.

The identification of the *hpa* gene encoding for heparanase enzyme will enable the production of a recombinant enzyme in heterologous expression systems. Availability of the recombinant protein will pave the way for solving the protein structure function relationship and will provide a tool for developing new inhibitors.

Viral Infection: The presence of heparan sulfate on cell surfaces have been shown to be the principal requirement for the binding of Herpes Simplex (33) and Dengue (33a) viruses to cells and for subsequent infection of the cells. Removal of the cell surface heparan sulfate by heparanase may

therefore abolish virus infection. In fact, treatment of cells with bacterial heparitinase (degrading heparan sulfate) or heparinase (degrading heparan) reduced the binding of two related animal herpes viruses to cells and rendered the cells at least partially resistant to virus infection (33). There 5 are some indications that the cell surface heparan sulfate is also involved in HIV infection (33b).

10 **Neurodegenerative diseases:** Heparan sulfate proteoglycans were identified in the prion protein amyloid plaques of Genstmann-Straussler Syndrome, Creutzfeldt-Jakob disease and Scrape (34). Heparanase may disintegrate these amyloid plaques which are also thought to play a role in the pathogenesis of Alzheimer's disease.

15 **Restenosis and Atherosclerosis:** Proliferation of arterial smooth muscle cells (SMCs) in response to endothelial injury and accumulation of cholesterol rich lipoproteins are basic events in the pathogenesis of atherosclerosis and restenosis (35). Apart from its involvement in SMC proliferation (i.e., low affinity receptors for heparin-binding growth factors), HS is also involved in lipoprotein binding, retention and uptake (36). It was demonstrated that HSPG and lipoprotein lipase participate in a novel 20 catabolic pathway that may allow substantial cellular and interstitial accumulation of cholesterol rich lipoproteins (32). The latter pathway is expected to be highly atherogenic by promoting accumulation of apoB and apoE rich lipoproteins (i.e. LDL, VLDL, chylomicrons), independent of feed back inhibition by the cellular sterol content. Removal of SMC HS by heparanase is therefore expected to inhibit both SMC proliferation and lipid 25 accumulation and thus may halt the progression of restenosis and atherosclerosis.

Gene therapy:

30 The ultimate goal in the management of inherited as well as acquired diseases is a rational therapy with the aim to eliminate the underlying biochemical defects associated with the disease rather than symptomatic treatment. Gene therapy is a promising candidate to meet these objectives. Initially it was developed for treatment of genetic disorders, however, the 35 consensus view today is that it offers the prospect of providing therapy for a variety of acquired diseases, including cancer, viral infections, vascular diseases and neurodegenerative disorders.

The gene-based therapeutic can act either intracellularly, affecting only the cells to which it is delivered, or extracellularly, using the recipient cells as local endogenous factories for the therapeutic product(s). The

application of gene therapy may follow any of the following strategies: (i) prophylactic gene therapy, such as using gene transfer to protect cells against viral infection; (ii) cytotoxic gene therapy, such as cancer therapy, where genes encode cytotoxic products to render the target cells vulnerable 5 to attack by the normal immune response; (iii) biochemical correction, primarily for the treatment of single gene defects, where a normal copy of the gene is added to the affected or other cells.

10 To allow efficient transfer of the therapeutic genes, a variety of gene delivery techniques have been developed based on viral and non-viral vector systems. The most widely used and most efficient systems for delivering genetic material into target cells are viral vectors. So far, 329 clinical studies (phase I, I/II and II) with over 2,500 patients have been initiated Worldwide since 1989 (50).

15 The approach of gene addition pose serious barriers. The expression of many genes is tightly regulated and context dependent, so achieving the correct balance and function of expression is challenging. The gene itself is often quite large, containing many exons and introns. The delivery vector is usually a virus, which can infect with a high efficiency but may, on the other hand, induce immunological response and consequently decreases 20 effectiveness, especially upon secondary administration. Most of the current expression vector-based gene therapy protocols fail to achieve clinically significant transgene expression required for treating genetic diseases. Apparently, it is difficult to deliver enough virus to the right cell type to elicit an effective and therapeutic effect (51)

25 Homologous recombination, which was initially considered to be of limited use for gene therapy because of its low frequency in mammalian cells, has recently emerged as a potential strategy for developing gene therapy. Different approaches have been used to study homologous recombination in mammalian cells; some involve DNA repair mechanisms. 30 These studies aimed at either gene disruption or gene correction and include RNA/DNA chimeric oligonucleotides, small or large homologous DNA fragments, or adeno-associated viral vectors. Most of these studies show a reasonable frequency of homologous recombination, which warrants further *in vivo* testing (52). Homologous recombination-based gene therapy has the 35 potential to develop into a powerful therapeutic modality for genetic diseases. It can offer permanent expression and normal regulation of corrected genes in appropriate cells or organs and probably can be used for treating dominantly inherited diseases such as polycystic kidney disease.

Genomic sequences function in regulation of gene expression:

The efficient expression of therapeutic genes in target cells or tissues is an important component of efficient and safe gene therapy. The expression of genes is driven by the promoter region upstream of the coding sequence, although regulation of expression may be supplemented by farther upstream or downstream DNA sequences or DNA in the introns of the gene. Since this important information is embedded in the DNA, the description of gene structure is crucial to the analysis of gene regulation. Characterization of cell specific or tissue specific promoters, as well as other tissue specific regulatory elements enables the use of such sequences to direct efficient cell specific, or developmental stage specific gene expression. This information provides the basis for targeting individual genes and for control of their expression by exogenous agents, such as drugs. Identification of transcription factors and other regulatory proteins required for proper gene expression will point at new potential targets for modulating gene expression, when so desired or required.

Efficient expression of many mammalian genes depends on the presence of at least one intron. The expression of mouse thymidylate synthase (TS) gene, for example, is greatly influenced by intron sequences. The addition of almost any of the introns from the mouse TS gene to an intronless TS minigene leads to a large increase in expression (42). The involvement of intron 1 in the regulation of expression was demonstrated for many other genes. In human factor IX (hFIX), intron 1 is able to increase the expression level about 3 fold more as compared to that of the hFIX cDNA (43). The expression enhancing activity of intron 1 is due to efficient functional splicing sequences, present in the precursor mRNA. By being efficiently assembled into spliceosome complexes, transcripts with splicing sequences may be better protected in the nucleus from random degradations, than those without such sequences (44).

A forward-inserted intron1-carrying hFIX expression cassette suggested to be useful for directed gene transfer, while for retroviral-mediated gene transfer system, reversely-inserted intron 1-carrying hFIX expression cassette was considered (43).

A highly conserved cis-acting sequence element was identified in the first intron of the mouse and rat c-Ha-ras, and in the first exon of Ha- and Ki-ras genes of human, mouse and rat. This cis-acting regulatory sequence confers strong transcription enhancer activity that is differentially modulated by steroid hormones in metastatic and nonmetastatic

subpopulations. Perturbations in the regulatory activities of such cis-acting sequences may play an important role in governing oncogenic potency of Ha-ras through transcriptional control mechanisms (45).

Intron sequences affect tissue specific, as well as inducible gene expression. A 182 bp intron 1 DNA segment of the mouse Col2a1 gene contains the necessary information to confer high-level, temporally correct, chondrocyte expression on a reporter gene in intact mouse embryos, while Col2a1 promoter sequences are dispensable for chondrocyte expression (46). In Col1A1 gene the intron plays little or no role in constitutive expression of collagen in the skin, and in cultured cells derived from the skin, however, in the lungs of young mice, intron deletion results in decrease of expression to less than 50 % (47).

A classical enhancer activity was shown in the 2 kb intron fragment in bovine beta-casein gene. The enhancer activity was largely dependent on the lactogenic hormones, especially prolactin. It was suggested that several elements in the intron-1 of the bovine beta-casein gene cooperatively interact not only with each other but also with its promoter for hormonal induction (48).

Identification and characterization of regulatory elements in genomic non-coding sequences, such as introns, provides a tool for designing and constructing novel vectors for tissue specific, hormone regulated or any other defined expression pattern, for gene therapy. Such an expression cassette was developed, utilizing regulatory elements from the human cytokeratin 18 (K18) gene, including 5' genomic sequences and one of its introns. This cassette efficiently expresses reporter genes, as well as the human cystic fibrosis transmembrane conductance regulator (CFTR) gene, in cultured lung epithelial cells (49).

Alternative splicing:

Alternative splicing of pre mRNA is a powerful and versatile regulatory mechanism that can effect quantitative control of gene expression and functional diversification of proteins. It contributes to major developmental decisions and also to a fine-tuning of gene function. Genetic and biochemical approaches have identified cis-acting regulatory elements and trans-acting factors that control alternative splicing of specific mRNAs. This mechanism results in the generation of variant isoforms of various proteins from a single gene. These include cell surface molecules such as CD44, receptors, cytokines such as VEGF and enzymes. Products of

alternatively spliced transcripts differ in their expression pattern, substrate specificity and other biological parameters.

5 The FGF receptor RNA undergoes alternative splicing which results in the production of several isoforms, which exhibit different ligand binding specificities. The alternative splicing is regulated in a cell specific manner (53).

10 Alternative spliced mRNAs are often correlated with malignancy. An increase in specific splice variant of tyrosinase was identified in murine melanomas (54). Multiple splicing variants of estrogen receptor are present 15 in individual human breast tumors. CD44 has various isoform, some are characteristic of malignant tissues.

15 Identification of tumor specific alternative splice variants provide new tool for cancer diagnostics. CD44 variants have been used for detection of malignancy in urine samples from patients with urothelial cancer by competitive RT-PCR (55). CD44 exon 6 was suggested as prognostic indicator of metastasis in breast cancer (56).

20 Different enzymes or polypeptides generated by alternative splicing may have different function or catalytic specificity. The identification and characterization of the enzyme forms, which are involved in pathological processes, is crucial for the design of appropriate and efficient drugs.

Modulation of gene expression – Antisense technology:

25 An antisense oligonucleotide (e.g., antisense oligodeoxyribonucleotide) may bind its target nucleic acid either by Watson-Crick base pairing or Hoogsteen and anti-Hoogsteen base pairing (64). According to the Watson-Crick base pairing, heterocyclic bases of the antisense oligonucleotide form hydrogen bonds with the heterocyclic bases 30 of target single-stranded nucleic acids (RNA or single-stranded DNA), whereas according to the Hoogsteen base pairing, the heterocyclic bases of the target nucleic acid are double-stranded DNA, wherein a third strand is accommodated in the major groove of the B-form DNA duplex by Hoogsteen and anti-Hoogsteen base pairing to form a triple helix structure.

35 According to both the Watson-Crick and the Hoogsteen base pairing models, antisense oligonucleotides have the potential to regulate gene expression and to disrupt the essential functions of the nucleic acids in cells. Therefore, antisense oligonucleotides have possible uses in modulating a wide range of diseases in which gene expression is altered.

Since the development of effective methods for chemically synthesizing oligonucleotides, these molecules have been extensively used

in biochemistry and biological research and have the potential use in medicine, since carefully devised oligonucleotides can be used to control gene expression by regulating levels of transcription, transcripts and/or translation.

5 Oligodeoxyribonucleotides as long as 100 base pairs (bp) are routinely synthesized by solid phase methods using commercially available, fully automated synthesis machines. The chemical synthesis of oligoribonucleotides, however, is far less routine. Oligoribonucleotides are also much less stable than oligodeoxyribonucleotides, a fact which has
10 contributed to the more prevalent use of oligodeoxyribonucleotides in medical and biological research, directed at, for example, the regulation of transcription or translation levels.

15 Gene expression involves few distinct and well regulated steps. The first major step of gene expression involves transcription of a messenger RNA (mRNA) which is an RNA sequence complementary to the antisense (i.e., -) DNA strand, or, in other words, identical in sequence to the DNA sense (i.e., +) strand, composing the gene. In eukaryotes, transcription occurs in the cell nucleus.

20 The second major step of gene expression involves translation of a protein (e.g., enzymes, structural proteins, secreted proteins, gene expression factors, etc.) in which the mRNA interacts with ribosomal RNA complexes (ribosomes) and amino acid activated transfer RNAs (tRNAs) to direct the synthesis of the protein coded for by the mRNA sequence.

25 Initiation of transcription requires specific recognition of a promoter DNA sequence located upstream to the coding sequence of a gene by an RNA-synthesizing enzyme -- RNA polymerase. This recognition is preceded by sequence-specific binding of one or more transcription factors to the promoter sequence. Additional proteins which bind at or close to the promoter sequence may trans upregulate transcription via cis elements
30 known as enhancer sequences. Other proteins which bind to or close to the promoter, but whose binding prohibits the action of RNA polymerase, are known as repressors.

35 There are also evidence that in some cases gene expression is downregulated by endogenous antisense RNA repressors that bind a complementary mRNA transcript and thereby prevent its translation into a functional protein.

Thus, gene expression is typically upregulated by transcription factors and enhancers and downregulated by repressors.

5 However, in many disease situations gene expression is impaired. In many cases, such as different types of cancer, for various reasons the expression of a specific endogenous or exogenous (e.g., of a pathogen such as a virus) gene is upregulated. Furthermore, in infectious diseases caused by pathogens such as parasites, bacteria or viruses, the disease progression depends on expression of the pathogen genes, this phenomenon may also be considered as far as the patient is concerned as upregulation of exogenous genes.

10 Most conventional drugs function by interaction with and modulation of one or more targeted endogenous or exogenous proteins, e.g., enzymes. Such drugs, however, typically are not specific for targeted proteins but interact with other proteins as well. Thus, a relatively large dose of drug must be used to effectively modulate a targeted protein.

15 Typical daily doses of drugs are from 10^{-5} - 10^{-1} millimoles per kilogram of body weight or 10^{-3} - 10 millimoles for a 100 kilogram person. If this modulation instead could be effected by interaction with and inactivation of mRNA, a dramatic reduction in the necessary amount of drug could likely be achieved, along with a corresponding reduction in side effects. Further reductions could be effected if such interaction could be 20 rendered site-specific. Given that a functioning gene continually produces mRNA, it would thus be even more advantageous if gene transcription could be arrested in its entirety.

Given these facts, it would be advantageous if gene expression could be arrested or downmodulated at the transcription level.

25 The ability of chemically synthesizing oligonucleotides and analogs thereof having a selected predetermined sequence offers means for downmodulating gene expression. Three types of gene expression modulation strategies may be considered.

30 At the transcription level, antisense or sense oligonucleotides or analogs that bind to the genomic DNA by strand displacement or the formation of a triple helix, may prevent transcription (64).

35 At the transcript level, antisense oligonucleotides or analogs that bind target mRNA molecules lead to the enzymatic cleavage of the hybrid by intracellular RNase H (65). In this case, by hybridizing to the targeted mRNA, the oligonucleotides or oligonucleotide analogs provide a duplex hybrid recognized and destroyed by the RNase H enzyme. Alternatively, such hybrid formation may lead to interference with correct splicing (66).

As a result, in both cases, the number of the target mRNA intact transcripts ready for translation is reduced or eliminated.

At the translation level, antisense oligonucleotides or analogs that bind target mRNA molecules prevent, by steric hindrance, binding of 5 essential translation factors (ribosomes), to the target mRNA, a phenomenon known in the art as hybridization arrest, disabling the translation of such mRNAs (67).

Thus, antisense sequences, which as described hereinabove may arrest the expression of any endogenous and/or exogenous gene depending 10 on their specific sequence, attracted much attention by scientists and pharmacologists who were devoted at developing the antisense approach into a new pharmacological tool (68).

For example, several antisense oligonucleotides have been shown to arrest hematopoietic cell proliferation (69), growth (70), entry into the S 15 phase of the cell cycle (71), reduced survival (72) and prevent receptor mediated responses (73). For use of antisense oligonucleotides as antiviral agents the reader is referred to reference 74.

For efficient *in vivo* inhibition of gene expression using antisense oligonucleotides or analogs, the oligonucleotides or analogs must fulfill the 20 following requirements (i) sufficient specificity in binding to the target sequence; (ii) solubility in water; (iii) stability against intra- and extracellular nucleases; (iv) capability of penetration through the cell membrane; and (v) when used to treat an organism, low toxicity.

Unmodified oligonucleotides are impractical for use as antisense 25 sequences since they have short *in vivo* half-lives, during which they are degraded rapidly by nucleases. Furthermore, they are difficult to prepare in more than milligram quantities. In addition, such oligonucleotides are poor cell membrane penetrators (75).

Thus it is apparent that in order to meet all the above listed 30 requirements, oligonucleotide analogs need to be devised in a suitable manner. Therefore, an extensive search for modified oligonucleotides has been initiated.

For example, problems arising in connection with double-stranded DNA (dsDNA) recognition through triple helix formation have been 35 diminished by a clever "switch back" chemical linking, whereby a sequence of polypurine on one strand is recognized, and by "switching back", a homopurine sequence on the other strand can be recognized. Also, good

helix formation has been obtained by using artificial bases, thereby improving binding conditions with regard to ionic strength and pH.

In addition, in order to improve half-life as well as membrane penetration, a large number of variations in polynucleotide backbones have been done, nevertheless with little success.

Oligonucleotides can be modified either in the base, the sugar or the phosphate moiety. These modifications include, for example, the use of methylphosphonates, monothiophosphates, dithiophosphates, phosphoramidates, phosphate esters, bridged phosphorothioates, bridged phosphoramidates, bridged methylenephosphonates, dephospho internucleotide analogs with siloxane bridges, carbonate bridges, carboxymethyl ester bridges, carbonate bridges, carboxymethyl ester bridges, acetamide bridges, carbamate bridges, thioether bridges, sulfoxyl bridges, sulfone bridges, various "plastic" DNAs, α -anomeric bridges and borane derivatives. For further details the reader is referred to reference 76.

International patent application WO 89/12060 discloses various building blocks for synthesizing oligonucleotide analogs, as well as oligonucleotide analogs formed by joining such building blocks in a defined sequence. The building blocks may be either "rigid" (i.e., containing a ring structure) or "flexible" (i.e., lacking a ring structure). In both cases, the building blocks contain a hydroxy group and a mercapto group, through which the building blocks are said to join to form oligonucleotide analogs. The linking moiety in the oligonucleotide analogs is selected from the group consisting of sulfide (-S-), sulfoxide (-SO-), and sulfone (-SO₂-). However, the application provides no data supporting the specific binding of an oligonucleotide analog to a target oligonucleotide.

International patent application WO 92/20702 describes an acyclic oligonucleotide which includes a peptide backbone on which any selected chemical nucleobases or analogs are stringed and serve as coding characters as they do in natural DNA or RNA. These new compounds, known as peptide nucleic acids (PNAs), are not only more stable in cells than their natural counterparts, but also bind natural DNA and RNA 50 to 100 times more tightly than the natural nucleic acids cling to each other (77). PNA oligomers can be synthesized from the four protected monomers containing thymine, cytosine, adenine and guanine by Merrifield solid-phase peptide synthesis. In order to increase solubility in water and to prevent aggregation, a lysine amide group is placed at the C-terminal.

Thus, antisense technology requires pairing of messenger RNA with an oligonucleotide to form a double helix that inhibits translation. The concept of antisense-mediated gene therapy was already introduced in 1978 for cancer therapy. This approach was based on certain genes that are 5 crucial in cell division and growth of cancer cells. Synthetic fragments of genetic substance DNA can achieve this goal. Such molecules bind to the targeted gene molecules in RNA of tumor cells, thereby inhibiting the translation of the genes and resulting in dysfunctional growth of these cells. Other mechanisms has also been proposed. These strategies have been 10 used, with some success in treatment of cancers, as well as other illnesses, including viral and other infectious diseases. Antisense oligonucleotides are typically synthesized in lengths of 13-30 nucleotides. The life span of oligonucleotide molecules in blood is rather short. Thus, they have to be chemically modified to prevent destruction by ubiquitous nucleases present 15 in the body. Phosphorothioates are very widely used modification in antisense oligonucleotide ongoing clinical trials (57). A new generation of antisense molecules consist of hybrid antisense oligonucleotide with a central portion of synthetic DNA while four bases on each end have been modified with 2'O-methyl ribose to resemble RNA. In preclinical studies in 20 laboratory animals, such compounds have demonstrated greater stability to metabolism in body tissues and an improved safety profile when compared with the first-generation unmodified phosphorothioate (Hybridon Inc. news). Doses of other nucleotide analogs have also been tested in antisense technology.

25 RNA oligonucleotides may also be used for antisense inhibition as they form a stable RNA-RNA duplex with the target, suggesting efficient inhibition. However, due to their low stability RNA oligonucleotides are typically expressed inside the cells using vectors designed for this purpose. This approach is favored when attempting to target a mRNA that encodes 30 an abundant and long-lived protein (57).

Recent scientific publications have validated the efficacy of antisense 35 compounds in animal models of hepatitis, cancers, coronary artery restenosis and other diseases. The first antisense drug was recently approved by the FDA. This drug Fomivirsen, developed by Isis, is indicated for local treatment of cytomegalovirus in patients with AIDS who are intolerant of or have a contraindication to other treatments for CMV retinitis or who were insufficiently responsive to previous treatments for CMV retinitis (Pharmacotherapy News Network).

Several antisense compounds are now in clinical trials in the United States. These include locally administered antivirals, systemic cancer therapeutics. Antisense therapeutics has the potential to treat many life-threatening diseases with a number of advantages over traditional drugs.

5 Traditional drugs intervene after a disease-causing protein is formed. Antisense therapeutics, however, block mRNA transcription/translation and intervene before a protein is formed, and since antisense therapeutics target only one specific mRNA, they should be more effective with fewer side effects than current protein-inhibiting therapy.

10 A second option for disrupting gene expression at the level of transcription uses synthetic oligonucleotides capable of hybridizing with double stranded DNA. A triple helix is formed. Such oligonucleotides may prevent binding of transcription factors to the gene's promoter and therefore inhibit transcription. Alternatively, they may prevent duplex unwinding
15 and, therefore, transcription of genes within the triple helical structure.

20 Another approach is the use of specific nucleic acid sequences to act as decoys for transcription factors. Since transcription factors bind specific DNA sequences it is possible to synthesize oligonucleotides that will effectively compete with the native DNA sequences for available transcription factors *in vivo*. This approach requires the identification of
25 gene specific transcription factor (57).

25 Indirect inhibition of gene expression was demonstrated for matrix metalloproteinase genes (MMP-1, -3, and -9), which are associated with invasive potential of human cancer cells. E1AF is a transcription activator of MMP genes. Expression of E1AF antisense RNA in HSC3AS cells showed decrease in mRNA and protein levels of MMP-1, -3, and -9. Moreover, HSC3AS showed lower invasive potential *in vitro* and *in vivo*. These results imply that transfection of antisense inhibits tumor invasion by down-regulating MMP genes (58).

30 **Ribozymes:**

35 Ribozymes are being increasingly used for the sequence-specific inhibition of gene expression by the cleavage of mRNAs encoding proteins of interest. The possibility of designing ribozymes to cleave any specific target RNA has rendered them valuable tools in both basic research and therapeutic applications. In the therapeutics area, ribozymes have been exploited to target viral RNAs in infectious diseases, dominant oncogenes in cancers and specific somatic mutations in genetic disorders. Most notably, several ribozyme gene therapy protocols for HIV patients are

already in Phase 1 trials (62). More recently, ribozymes have been used for transgenic animal research, gene target validation and pathway elucidation. Several ribozymes are in various stages of clinical trials. ANGIOZYME was the first chemically synthesized ribozyme to be studied in human 5 clinical trials. ANGIOZYME specifically inhibits formation of the VEGF-r (Vascular Endothelial Growth Factor receptor), a key component in the angiogenesis pathway. Ribozyme Pharmaceuticals, Inc., as well as other firms have demonstrated the importance of anti-angiogenesis therapeutics in animal models. HEPTAZYME, a ribozyme designed to selectively destroy 10 Hepatitis C Virus (HCV) RNA, was found effective in decreasing Hepatitis C viral RNA in cell culture assays (Ribozyme Pharmaceuticals, Incorporated - WEB home page).

Gene disruption in animal models:

The emergence of gene inactivation by homologous recombination 15 methodology in embryonic stem cells has revolutionized the field of mouse genetics. The availability of a rapidly growing number of mouse null mutants has represented an invaluable source of knowledge on mammalian development, cellular biology and physiology, and has provided many models for human inherited diseases. Animal models are required for an 20 effective drug delivery development program and evaluation of gene therapy approach. The improvement of the original knockout strategy, as well as exploitation of exogenous enzymatic systems that are active in the recombination process, has been considerably extended the range of genetic 25 manipulations that can be produced. Additional methods have been developed to provide versatile research tools: Double replacement method, sequential gene targeting, conditional cell type specific gene targeting, single copy integration method, inducible gene targeting, gene disruption by viral delivery, replacing one gene with another, the so called knock-in method and the induction of specific balanced chromosomal translocation. 30 It is now possible to introduce a point mutation as a unique change in the entire genome, therefore allowing very fine dissection of gene function *in vivo*. Furthermore, the advent of methods allowing conditional gene targeting opens the way for analysis of consequence of a particular mutation 35 in a defined organ and at a specific time during the life of the experimental animal (59).

DNA vaccination:

Observations in the early 1990s that plasmid DNA could directly transfet animal cells *in vivo* sparked exploration of the use of DNA

plasmids to induce immune response by direct injection into animal of DNA encoding antigenic protein. When a DNA vaccine plasmid enters the eukaryotic cell, the protein it encodes is transcribed and translated within the cell. In the case of pathogens, these proteins are presented to the immune system in their native form, mimicking the presentation of antigens during a natural infection. DNA vaccination is particularly useful for the induction of T cell activation. It was applied for viral and bacterial infectious diseases, as well as for allergy and for cancer. The central hypothesis behind active specific immunotherapy for cancer is that tumor cells express unique antigens that should stimulate the immune system. The first DNA vaccine against tumor was carcino-embryonic antigen (CEA). DNA vaccinated animals expressed immunoprotection and immunotherapy of human CEA-expressing syngeneic mouse colon and breast carcinoma (61). In a mouse model of neuroblastoma, DNA immunization with HuD resulted in tumor growth inhibition with no neurological disease (60). Immunity to the brown locus protein, gp⁷⁵ tyrosinase-related protein-1, associated with melanoma, was investigated in a syngeneic mouse model. Priming with human gp75 DNA broke tolerance to mouse gp75. Immunity against mouse gp75 provided significant tumor protection (60).

20 Glycosyl hydrolases:

Glycosyl hydrolases are a widespread group of enzymes that hydrolyze the α -glycosidic bond between two or more carbohydrates or between a carbohydrate and a noncarbohydrate moiety. The enzymatic hydrolysis of glycosidic bond occurs by using major one or two mechanisms leading to overall retention or inversion of the anomeric configuration. In both mechanisms catalysis involves two residues: a proton donor and a nucleophile. Glycosyl hydrolyses have been classified into 58 families based on amino acid similarities. The glycosyl hydrolyses from families 1, 2, 5, 10, 17, 30, 35, 39 and 42 act on a large variety of substrates, however, they all hydrolyze the glycosidic bond in a general acid catalysis mechanism, with retention of the anomeric configuration. The mechanism involves two glutamic acid residues, which are the proton donors and the nucleophile, with an asparagine always preceding the proton donor. Analyses of a set of known 3D structures from this group revealed that their catalytic domains, despite the low level of sequence identity, adopt a similar (α/β)₈ fold with the proton donor and the nucleophile located at the C-terminal ends of strands β 4 and β 7, respectively. Mutations

in the functional conserved amino acids of lysosomal glycosyl hydrolases were identified in lysosomal storage diseases.

Lysosomal glycosyl hydrolases including β -glucuronidase, β -mannosidase, β -glucocerebrosidase, β -galactosidase and α -L iduronidase, are all exo-glycosyl hydrolases, belong to the GH-A clan and share a similar catalytic site. However, many endo-glucanases from various organisms, such as bacterial and fungal xylanases and cellulases share this catalytic domain.

Genomic sequence of hpa gene and its implications:

It is well established that heparanase activity is correlated with cancer metastasis. This correlation was demonstrated at the level of enzymatic activity as well as the levels of protein and *hpa* cDNA expression in highly metastatic cancer cells as compared with non-metastatic cells. As such, inhibition of heparanase activity is desirable, and has been attempted by several means. The genomic region, encoding the *hpa* gene and the surrounding, provides a new powerful tool for regulation of heparanase activity at the level of gene expression. Regulatory sequences may reside in noncoding regions both upstream and downstream the transcribed region as well as in intron sequences. A DNA sequence upstream of the transcription start site contains the promoter region and potential regulatory elements. Regulatory factors, which interact with the promoter region may be identified and be used as potential drugs for inhibition of cancer, metastasis and inflammation. The promoter region can be used to screen for inhibitors of heparanase gene expression. Furthermore, the *hpa* promoter can be used to direct cell specific, particularly cancer cell specific, expression of foreign genes, such as cytotoxic or apoptotic genes, in order to specifically destroy cancer cells.

Cancer and yet unknown related genetic disorders may involve rearrangements and mutations in the heparanase gene, either in coding or non-coding regions. Such mutations may affect expression level or enzymatic activity. The genomic sequence of *hpa* enables the amplification of specific genomic DNA fragments, identification and diagnosis of mutations.

There is thus a widely recognized need for, and it would be highly advantageous to have genomic, cDNA and composite polynucleotides encoding a polypeptide having heparanase activity, vectors including same, genetically modified cells expressing heparanase and a recombinant protein

having heparanase activity, as well as antisense oligonucleotides, constructs and ribozymes which can be used for down regulation heparanase activity.

SUMMARY OF THE INVENTION

5 Cloning of the human *hpa* gene which encodes heparanase, and expression of recombinant heparanase by transfected host cells is reported herein, as well as downregulation of heparanase activity by antisense technology.

10 A purified preparation of heparanase isolated from human hepatoma cells was subjected to tryptic digestion and microsequencing. The YGPDVGVQPR (SEQ ID NO:8) sequence revealed was used to screen EST databases for homology to the corresponding back translated DNA sequence. Two closely related EST sequences were identified and were thereafter found to be identical. Both clones contained an insert of 1020 bp 15 which included an open reading frame of 973 bp followed by a 27 bp of 3' untranslated region and a Poly A tail. Translation start site was not identified.

20 Cloning of the missing 5' end of *hpa* was performed by PCR amplification of DNA from placenta Marathon RACE cDNA composite using primers selected according to the EST clones sequence and the linkers of the composite. A 900 bp PCR fragment, partially overlapping with the identified 3' encoding EST clones was obtained. The joined cDNA 25 fragment (*hpa*), 1721 bp long (SEQ ID NO:9), contained an open reading frame which encodes a polypeptide of 543 amino acids (SEQ ID NO:10) with a calculated molecular weight of 61,192 daltons.

Cloning an extended 5' sequence was enabled from the human SK-hep1 cell line by PCR amplification using the Marathon RACE. The 5' extended sequence of the SK-hep1 *hpa* cDNA was assembled with the sequence of the *hpa* cDNA isolated from human placenta (SEQ ID NO:9). 30 The assembled sequence contained an open reading frame, SEQ ID NOs: 13 and 15, which encodes, as shown in SEQ ID NOs:14 and 15, a polypeptide of 592 amino acids with a calculated molecular weight of 66,407 daltons.

35 The ability of the *hpa* gene product to catalyze degradation of heparan sulfate in an *in vitro* assay was examined by expressing the entire open reading frame of *hpa* in insect cells, using the Baculovirus expression system. Extracts and conditioned media of cells infected with virus containing the *hpa* gene, demonstrated a high level of heparan sulfate degradation activity both towards soluble ECM-derived HSPG and intact

ECM. This degradation activity was inhibited by heparin, which is another substrate of heparanase. Cells infected with a similar construct containing no *hpa* gene had no such activity, nor did non-infected cells. The ability of heparanase expressed from the extended 5' clone towards heparin was demonstrated in a mammalian expression system.

The expression pattern of *hpa* RNA in various tissues and cell lines was investigated using RT-PCR. It was found to be expressed only in tissues and cells previously known to have heparanase activity.

A panel of monochromosomal human/CHO and human/mouse somatic cell hybrids was used to localize the human heparanase gene to human chromosome 4. The newly isolated heparanase sequence can be used to identify a chromosome region harboring a human heparanase gene in a chromosome spread.

A human genomic library was screened and the human locus harboring the heparanase gene isolated, sequenced and characterized. Alternatively spliced heparanase mRNAs were identified and characterized. The human heparanase promoter has been isolated, identified and positively tested for activity. The mouse heparanase promoter has been isolated and identified as well. Antisense heparanase constructs were prepared and their influence on cells *in vitro* tested. A predicted heparanase active site was identified. And finally, the presence of sequences hybridizing with human heparanase sequences was demonstrated for a variety of mammals and for an avian.

According to one aspect of the present invention there is provided an isolated nucleic acid comprising a genomic, complementary or composite polynucleotide sequence encoding a polypeptide having heparanase catalytic activity.

According to further features in preferred embodiments of the invention described below, the polynucleotide or a portion thereof is hybridizable with SEQ ID NOs: 9, 13, 42, 43 or a portion thereof at 68 °C in 6 x SSC, 1 % SDS, 5 x Denharts, 10 % dextran sulfate, 100 µg/ml salmon sperm DNA, and ³²P labeled probe and wash at 68 °C with 3 x SSC and 0.1 % SDS.

According to still further features in the described preferred embodiments the polynucleotide or a portion thereof is at least 60 % identical with SEQ ID NOs: 9, 13, 42, 43 or portions thereof as determined using the Bestfit procedure of the DNA sequence analysis software package

developed by the Genetic Computer Group (GCG) at the university of Wisconsin (gap creation penalty - 12, gap extension penalty - 4).

According to still further features in the described preferred embodiments the polypeptide is as set forth in SEQ ID NOs:10, 14, 44 or portions thereof.

According to still further features in the described preferred embodiments the polypeptide is at least 60 % homologous to SEQ ID NOs:10, 14, 44 or portions thereof as determined with the Smith-Waterman algorithm, using the Bioaccelerator platform developed by Compugene (gapop: 10.0, gapext: 0.5, matrix: blosum62).

According to additional aspects of the present invention there are provided a nucleic acid construct (vector) comprising the isolated nucleic acid described herein and a host cell comprising the construct.

According to a further aspect of the present invention there is provided an antisense oligonucleotide comprising a polynucleotide or a polynucleotide analog of at least 10 bases being hybridizable *in vivo*, under physiological conditions, with a portion of a polynucleotide strand encoding a polypeptide having heparanase catalytic activity.

According to an additional aspect of the present invention there is provided a method of *in vivo* downregulating heparanase activity comprising the step of *in vivo* administering the antisense oligonucleotide herein described.

According to yet an additional aspect of the present invention there is provided a pharmaceutical composition comprising the antisense oligonucleotide herein described and a pharmaceutically acceptable carrier.

According to still an additional aspect of the present invention there is provided a ribozyme comprising the antisense oligonucleotide described herein and a ribozyme sequence.

According to a further aspect of the present invention there is provided an antisense nucleic acid construct comprising a promoter sequence and a polynucleotide sequence directing the synthesis of an antisense RNA sequence of at least 10 bases being hybridizable *in vivo*, under physiological conditions, with a portion of a polynucleotide strand encoding a polypeptide having heparanase catalytic activity.

According to further features in preferred embodiments of the invention described below, the polynucleotide strand encoding the polypeptide having heparanase catalytic activity is as set forth in SEQ ID NOs: 9, 13, 42 or 43.

According to still further features in the described preferred embodiments the polypeptide having heparanase catalytic activity is as set forth in SEQ ID NOs: 10, 14 or 44.

According to still a further aspect of the present invention there is provided a method of *in vivo* downregulating heparanase activity comprising the step of *in vivo* administering the antisense nucleic acid construct herein described.

According to yet a further aspect of the present invention there is provided a pharmaceutical composition comprising the antisense nucleic acid construct herein described and a pharmaceutically acceptable carrier.

According to a further aspect of the present invention there is provided a nucleic acid construct comprising a polynucleotide sequence functioning as a promoter, the polynucleotide sequence is derived from SEQ ID NO:42 and includes at least nucleotides 2535-2635 thereof or from SEQ ID NO:43 and includes at least nucleotides 320-420.

According to a further aspect of the present invention there is provided a method of expressing a polynucleotide sequence comprising the step of ligating the polynucleotide sequence into the nucleic acid construct described above, downstream of the polynucleotide sequence derived from SEQ ID NOs:42 or 43.

According to a further aspect of the present invention there is provided a recombinant protein comprising a polypeptide having heparanase catalytic activity.

According to further features in preferred embodiments of the invention described below, the polypeptide includes at least a portion of SEQ ID NOs:10, 14 or 44.

According to still further features in the described preferred embodiments the protein is encoded by a polynucleotide hybridizable with SEQ ID NOs: 9, 13, 42, 43 or a portion thereof at 68 °C in 6 x SSC, 1 % SDS, 5 x Denharts, 10 % dextran sulfate, 100 µg/ml salmon sperm DNA, and ³²P labeled probe and wash at 68 °C with 3 x SSC and 0.1 % SDS.

According to still further features in the described preferred embodiments the protein is encoded by a polynucleotide at least 60 % identical with SEQ ID NOs: 9, 13, 42, 43 or portions thereof as determined using the Bestfit procedure of the DNA sequence analysis software package developed by the Genetic Computer Group (GCG) at the university of Wisconsin (gap creation penalty - 12, gap extension penalty - 4).

According to a further aspect of the present invention there is provided a pharmaceutical composition comprising, as an active ingredient, the recombinant protein herein described.

According to a further aspect of the present invention there is provided a method of identifying a chromosome region harboring a heparanase gene in a chromosome spread comprising the steps of (a) hybridizing the chromosome spread with a tagged polynucleotide probe encoding heparanase; (b) washing the chromosome spread, thereby removing excess of non-hybridized probe; and (c) searching for signals associated with the hybridized tagged polynucleotide probe, wherein detected signals being indicative of a chromosome region harboring a heparanase gene.

According to a further aspect of the present invention there is provided a method of *in vivo* eliciting anti-heparanase antibodies comprising the steps of administering a nucleic acid construct including a polynucleotide segment corresponding to at least a portion of SEQ ID NOs:9, 13 or 43 and a promoter for directing the expression of said polynucleotide segment *in vivo*. Accordingly, there is provided also a DNA vaccine for *in vivo* eliciting anti-heparanase antibodies comprising a nucleic acid construct including a polynucleotide segment corresponding to at least a portion of SEQ ID NOs:9, 13 or 43 and a promoter for directing the expression of said polynucleotide segment *in vivo*.

The present invention can be used to develop new drugs to inhibit tumor cell metastasis, inflammation and autoimmunity. The identification of the *hpa* gene encoding for heparanase enzyme enables the production of a recombinant enzyme in heterologous expression systems. Additional features, advantages, uses and applications of the present invention in biological science and in diagnostic and therapeutic medicine are described hereinafter.

30

BRIEF DESCRIPTION OF THE DRAWINGS

The invention herein described, by way of example only, with reference to the accompanying drawings, wherein:

FIG. 1 presents nucleotide sequence and deduced amino acid sequence of *hpa* cDNA. A single nucleotide difference at position 799 (A to T) between the EST (Expressed Sequence Tag) and the PCR amplified cDNA (reverse transcribed RNA) and the resulting amino acid substitution (Tyr to Phe) are indicated above and below the substituted unit,

respectively. Cysteine residues and the poly adenylation consensus sequence are underlined. The asterisk denotes the stop codon TGA.

FIG. 2 demonstrates degradation of soluble sulfate labeled HSPG substrate by lysates of High Five cells infected with pF_{hpa2} virus. Lysates of High Five cells that were infected with pF_{hpa2} virus (●) or control pF2 virus (□) were incubated (18 h, 37 °C) with sulfate labeled ECM-derived soluble HSPG (peak I). The incubation medium was then subjected to gel filtration on Sepharose 6B. Low molecular weight HS degradation fragments (peak II) were produced only during incubation with the pF_{hpa2} infected cells, but there was no degradation of the HSPG substrate (◊) by lysates of pF2 infected cells.

FIGs. 3a-b demonstrate degradation of soluble sulfate labeled HSPG substrate by the culture medium of pF_{hpa2} and pF_{hpa4} infected cells. Culture media of High Five cells infected with pF_{hpa2} (3a) or pF_{hpa4} (3b) viruses (●), or with control viruses (□) were incubated (18 h, 37 °C) with sulfate labeled ECM-derived soluble HSPG (peak I, ◊). The incubation media were then subjected to gel filtration on Sepharose 6B. Low molecular weight HS degradation fragments (peak II) were produced only during incubation with the *hpa* gene containing viruses. There was no degradation of the HSPG substrate by the culture medium of cells infected with control viruses.

FIG. 4 presents size fractionation of heparanase activity expressed by pF_{hpa2} infected cells. Culture medium of pF_{hpa2} infected High Five cells was applied onto a 50 kDa cut-off membrane. Heparanase activity (conversion of the peak I substrate, (◊) into peak II HS degradation fragments) was found in the high (> 50 kDa) (●), but not low (< 50 kDa) (○) molecular weight compartment.

FIGs. 5a-b demonstrate the effect of heparin on heparanase activity expressed by pF_{hpa2} and pF_{hpa4} infected High Five cells. Culture media of pF_{hpa2} (5a) and pF_{hpa4} (5b) infected High Five cells were incubated (18 h, 37 °C) with sulfate labeled ECM-derived soluble HSPG (peak I, ◊) in the absence (●) or presence (Δ) of 10 µg/ml heparin. Production of low molecular weight HS degradation fragments was completely abolished in the presence of heparin, a potent inhibitor of heparanase activity (6, 7).

FIGs. 6a-b demonstrate degradation of sulfate labeled intact ECM by virus infected High Five and Sf21 cells. High Five (6a) and Sf21 (6b) cells were plated on sulfate labeled ECM and infected (48 h, 28 °C) with pF_{hpa4} (●) or control pF1 (□) viruses. Control non-infected Sf21 cells (R) were

plated on the labeled ECM as well. The pH of the cultured medium was adjusted to 6.0 - 6.2 followed by 24 h incubation at 37 °C. Sulfate labeled material released into the incubation medium was analyzed by gel filtration on Sepharose 6B. HS degradation fragments were produced only by cells infected with the *hpa* containing virus.

FIG. 7a-b demonstrate degradation of sulfate labeled intact ECM by virus infected cells. High Five (7a) and Sf21 (7b) cells were plated on sulfate labeled ECM and infected (48 h, 28 °C) with pF*hpa4* (●) or control pF1 (□) viruses. Control non-infected Sf21 cells (R) were plated on labeled ECM as well. The pH of the cultured medium was adjusted to 6.0 - 6.2, followed by 48 h incubation at 28 °C. Sulfate labeled degradation fragments released into the incubation medium was analyzed by gel filtration on Sepharose 6B. HS degradation fragments were produced only by cells infected with the *hpa* containing virus.

FIGs. 8a-b demonstrate degradation of sulfate labeled intact ECM by the culture medium of pF*hpa4* infected cells. Culture media of High Five (8a) and Sf21 (8b) cells that were infected with pF*hpa4* (●) or control pF1 (□) viruses were incubated (48 h, 37 °C, pH 6.0) with intact sulfate labeled ECM. The ECM was also incubated with the culture medium of control non-infected Sf21 cells (R). Sulfate labeled material released into the reaction mixture was subjected to gel filtration analysis. Heparanase activity was detected only in the culture medium of pF*hpa4* infected cells.

FIGs. 9a-b demonstrate the effect of heparin on heparanase activity in the culture medium of pF*hpa4* infected cells. Sulfate labeled ECM was incubated (24 h, 37 °C, pH 6.0) with culture medium of pF*hpa4* infected High Five (9a) and Sf21 (9b) cells in the absence (●) or presence (V) of 10 µg/ml heparin. Sulfate labeled material released into the incubation medium was subjected to gel filtration on Sepharose 6B. Heparanase activity (production of peak II HS degradation fragments) was completely inhibited in the presence of heparin.

FIGs. 10a-b demonstrate purification of recombinant heparanase on heparin-Sepharose. Culture medium of Sf21 cells infected with pF*hpa4* virus was subjected to heparin-Sepharose chromatography. Elution of fractions was performed with 0.35 - 2 M NaCl gradient (◊). Heparanase activity in the eluted fractions is demonstrated in Figure 10a (●). Fractions 15-28 were subjected to 15 % SDS-polyacrylamide gel electrophoresis followed by silver nitrate staining. A correlation is demonstrated between a

major protein band (MW ~ 63,000) in fractions 19 - 24 and heparanase activity.

FIGs. 11a-b demonstrate purification of recombinant heparanase on a Superdex 75 gel filtration column. Active fractions eluted from heparin-Sepharose (Figure 10a) were pooled, concentrated and applied onto Superdex 75 FPLC column. Fractions were collected and aliquots of each fraction were tested for heparanase activity (C, Figure 11a) and analyzed by SDS-polyacrylamide gel electrophoresis followed by silver nitrate staining (Figure 11b). A correlation is seen between the appearance of a major protein band (MW ~ 63,000) in fractions 4 - 7 and heparanase activity.

FIGs. 12a-e demonstrate expression of the *hpa* gene by RT-PCR with total RNA from human embryonal tissues (12a), human extra-embryonal tissues (12b) and cell lines from different origins (12c-e). RT-PCR products using *hpa* specific primers (I), primers for GAPDH housekeeping gene (II), and control reactions without reverse transcriptase demonstrating absence of genomic DNA or other contamination in RNA samples (III). M- DNA molecular weight marker VI (Boehringer Mannheim). For 12a: lane 1 - neutrophil cells (adult), lane 2 - muscle, lane 3 - thymus, lane 4 - heart, lane 5 - adrenal. For 12b: lane 1 - kidney, lane 2 - placenta (8 weeks), lane 3 - placenta (11 weeks), lanes 4-7 - mole (complete hydatidiform mole), lane 8 - cytotrophoblast cells (freshly isolated), lane 9 - cytotrophoblast cells (1.5 h *in vitro*), lane 10 - cytotrophoblast cells (6 h *in vitro*), lane 11 - cytotrophoblast cells (18 h *in vitro*), lane 12 - cytotrophoblast cells (48 h *in vitro*). For 12c: lane 1 - JAR bladder cell line, lane 2 - NCITT testicular tumor cell line, lane 3 - SW-480 human hepatoma cell line, lane 4 - HTR (cytotrophoblasts transformed by SV40), lane 5 - HPTLP-I hepatocellular carcinoma cell line, lane 6 - EJ-28 bladder carcinoma cell line. For 12d: lane 1 - SK-hep-1 human hepatoma cell line, lane 2 - DAMI human megakaryocytic cell line, lane 3 - DAMI cell line + PMA, lane 4 - CHRF cell line + PMA, lane 5 - CHRF cell line. For 12e: lane 1 - ABAE bovine aortic endothelial cells, lane 2 - 1063 human ovarian cell line, lane 3 - human breast carcinoma MDA435 cell line, lane 4 - human breast carcinoma MDA231 cell line.

FIG. 13 presents a comparison between nucleotide sequences of the human *hpa* and a mouse EST cDNA fragment (SEQ ID NO:12) which is 80 % homologous to the 3' end (starting at nucleotide 1066 of SEQ ID NO:9) of the human *hpa*. The aligned termination codons are underlined.

FIG. 14 demonstrates the chromosomal localization of the *hpa* gene. PCR products of DNA derived from somatic cell hybrids and of genomic DNA of hamster, mouse and human were separated on 0.7 % agarose gel following amplification with *hpa* specific primers. Lane 1 – Lambda DNA digested with *Bst*EII, lane 2 – no DNA control, lanes 3 – 29, PCR amplification products. Lanes 3-5 – human, mouse and hamster genomic DNA, respectively. Lanes 6-29, human monochromosomal somatic cell hybrids representing chromosomes 1-22 and X and Y, respectively. Lane 30 – Lambda DNA digested with *Bst*EII. An amplification product of approximately 2.8 Kb is observed only in lanes 5 and 9, representing human genomic DNA and DNA derived from cell hybrid carrying human chromosome 4, respectively. These results demonstrate that the *hpa* gene is localized in human chromosome 4.

FIG. 15 demonstrates the genomic exon-intron structure of the human *hpa* locus (top) and the relative positions of the lambda clones used as sequencing templates to sequence the locus (below). The vertical rectangles represent exons (E) and the horizontal lines therebetween represent introns (I), upstream (U) and downstream (D) regions. Continuous lines represent DNA fragments, which were used for sequence analysis. The discontinuous line in lambda 6 represent a region, which overlaps with lambda 8 and hence was not analyzed. The plasmid contains a PCR product, which bridges the gap between L3 and L6.

FIG. 16 presents the nucleotide sequence of the genomic region of the *hpa* gene. Exon sequences appear in upper case and intron sequences in lower case. The deduced amino acid sequence of the exons is printed below the nucleotide sequence. Two predicted transcription start sites are shown in bold.

FIG. 17 presents an alignment of the amino acid sequences of human heparanase, mouse and partial sequences of rat homologues. The human and the mouse sequences were determined by sequence analysis of the isolated cDNAs. The rat sequence is derived from two different EST clones, which represent two different regions (5' and 3') of the rat *hpa* cDNA. The human sequence and the amino acids in the mouse and rat homologues, which are identical to the human sequence, appear in bold.

FIG. 18 presents a heparanase Zoo blot. Ten micrograms of genomic DNA from various sources were digested with *Eco*RI and separated on 0.7 % agarose – TBE gel. Following electrophoresis, the gel was treated with HCl and then with NaOH and the DNA fragments were downward

transferred to a nylon membrane (Hybond N+, Amersham) with 0.4 N NaOH. The membrane was hybridized with a 1.6 Kb DNA probe that contained the entire *hpa* cDNA. Lane order: H – Human; M – Mouse; Rt – Rat; P – Pig; Cw – Cow; Hr – Horse; S – Sheep; Rb – Rabbit; D – Dog; Ch – Chicken; F – Fish. Size markers (Lambda *B*stII) are shown on the left

5 FIG. 19 demonstrates the secondary structure prediction for heparanase performed using the PHD server – Profile network Prediction Heidelberg. H – helix, E – extended (beta strand), The glutamic acid predicted as the proton donor is marked by asterisk and the possible
10 nucleophiles are underlined.

DESCRIPTION OF THE PREFERRED EMBODIMENTS

The present invention is of a polynucleotide or nucleic acid, referred to hereinbelow interchangeably as *hpa*, *hpa* cDNA or *hpa* gene or identified
15 by its SEQ ID NOs, encoding a polypeptide having heparanase activity, vectors or nucleic acid constructs including same and which are used for over-expression or antisense inhibition of heparanase, genetically modified cells expressing same, recombinant protein having heparanase activity, antisense oligonucleotides and ribozymes for heparanase modulation, and
20 heparanase promoter sequences which can be used to direct the expression of desired genes.

Before explaining at least one embodiment of the invention in detail, it is to be understood that the invention is not limited in its application to the details of construction and the arrangement of the components set forth in
25 the following description or illustrated in the drawings. The invention is capable of other embodiments or of being practiced or carried out in various ways. Also, it is to be understood that the phraseology and terminology employed herein is for the purpose of description and should not be regarded as limiting.

30 Cloning of the human and mouse *hpa* genes, cDNAs and genomic sequence (for human), encoding heparanase and expressing recombinant heparanase by transfected cells is reported herein. These are the first mammalian heparanase genes to be cloned.

35 A purified preparation of heparanase isolated from human hepatoma cells was subjected to tryptic digestion and microsequencing.

The YGPDVGQPR (SEQ ID NO:8) sequence revealed was used to screen EST databases for homology to the corresponding back translated

DNA sequences. Two closely related EST sequences were identified and were thereafter found to be identical.

Both clones contained an insert of 1020 bp which includes an open reading frame of 973 bp followed by a 3' untranslated region of 27 bp and a 5 Poly A tail, whereas a translation start site was not identified.

Cloning of the missing 5' end was performed by PCR amplification of DNA from placenta Marathon RACE cDNA composite using primers selected according to the EST clones sequence and the linkers of the composite.

10 A 900 bp PCR fragment, partially overlapping with the identified 3' encoding EST clones was obtained. The joined cDNA fragment (*hpa*), 1721 bp long (SEQ ID NO:9), contained an open reading frame which encodes, as shown in Figure 1 and SEQ ID NO:11, a polypeptide of 543 amino acids (SEQ ID NO:10) with a calculated molecular weight of 61,192 15 daltons.

20 A single nucleotide difference at position 799 (A to T) between the EST clones and the PCR amplified cDNA was observed. This difference results in a single amino acid substitution (Tyr to Phe) (Figure 1). Furthermore, the published EST sequences contained an unidentified 25 nucleotide, which following DNA sequencing of both the EST clones was resolved into two nucleotides (G and C at positions 1630 and 1631 in SEQ ID NO:9, respectively).

25 The ability of the *hpa* gene product to catalyze degradation of heparan sulfate in an *in vitro* assay was examined by expressing the entire open reading frame in insect cells, using the Baculovirus expression system.

30 Extracts and conditioned media of cells infected with virus containing the *hpa* gene, demonstrated a high level of heparan sulfate degradation activity both towards soluble ECM-derived HSPG and intact ECM, which was inhibited by heparin, while cells infected with a similar construct containing no *hpa* gene had no such activity, nor did non-infected cells.

The expression pattern of *hpa* RNA in various tissues and cell lines was investigated using RT-PCR. It was found to be expressed only in tissues and cells previously known to have heparanase activity.

35 Cloning an extended 5' sequence was enabled from the human SK-hep1 cell line by PCR amplification using the Marathon RACE. The 5' extended sequence of the SK-hep1 *hpa* cDNA was assembled with the sequence of the *hpa* cDNA isolated from human placenta (SEQ ID NO:9).

The assembled sequence contained an open reading frame, SEQ ID NOs: 13 and 15, which encodes, as shown in SEQ ID NOs:14 and 15, a polypeptide of 592 amino acids, with a calculated molecular weight of 66,407 daltons. This open reading frame was shown to direct the expression of catalytically active heparanase in a mammalian cell expression system. The expressed heparanase was detectable by anti heparanase antibodies in Western blot analysis.

A panel of monochromosomal human/CHO and human/mouse somatic cell hybrids was used to localize the human heparanase gene to human chromosome 4. The newly isolated heparanase sequence can therefore be used to identify a chromosome region harboring a human heparanase gene in a chromosome spread.

The *hpa* cDNA was then used as a probe to screen a human genomic library. Several phages were positive. These phages were analyzed and were found to cover most of the *hpa* locus, except for a small portion which was recovered by bridging PCR. The *hpa* locus covers about 50,000 bp. The *hpa* gene includes 12 exons separated by 11 introns.

RT-PCR performed on a variety of cells revealed alternatively spliced *hpa* transcripts.

The amino acid sequence of human heparanase was used to search for homologous sequences in the DNA and protein databases. Several human EST's were identified, as well as mouse sequences highly homologous to human heparanase. The following mouse EST's were identified AA177901, AA674378, AA67997, AA047943, AA690179, AI122034, all sharing an identical sequence and correspond to amino acids 336-543 of the human heparanase sequence. The entire mouse heparanase cDNA was cloned, based on the nucleotide sequence of the mouse EST's using Marathon cDNA libraries. The mouse and the human *hpa* genes share an average homology of 78 % between the nucleotide sequences and 81 % similarity between the deduced amino acid sequences. *hpa* homologous sequences from rat were also uncovered (EST's AI060284 and AI237828).

Homology search of heparanase amino acid sequence against the DNA and the protein databases and prediction of its protein secondary structure enabled to identify candidate amino acids that participate in the heparanase active site.

Expression of *hpa* antisense in mammalian cell lines resulted in about five fold decrease in the number of recoverable cells as compared to controls.

5 Human *Hpa* cDNA was shown to hybridize with genomic DNAs of a variety of mammalian species and with an avian.

The human and mouse *hpa* promoters were identified and the human promoter was tested positive in directing the expression of a reporter gene.

10 Thus, according to the present invention there is provided an isolated nucleic acid comprising a genomic, complementary or composite polynucleotide sequence encoding a polypeptide having heparanase catalytic activity.

15 The phrase "composite polynucleotide sequence" refers to a sequence which includes exon sequences required to encode the polypeptide having heparanase activity, as well as any number of intronal sequences. The intronal sequences can be of any source and typically will include conserved splicing signal sequences. Such intronal sequences may further include cis acting expression regulatory elements.

20 The term "heparanase catalytic activity" or its equivalent term "heparanase activity" both refer to a mammalian endoglycosidase hydrolyzing activity which is specific for heparan or heparan sulfate proteoglycan substrates, as opposed to the activity of bacterial enzymes (heparinase I, II and III) which degrade heparin or heparan sulfate by means of β -elimination (37).

25 According to a preferred embodiment of the present invention the polynucleotide or a portion thereof is hybridizable with SEQ ID NOs: 9, 13, 42, 43 or a portion thereof at 68 °C in 6 x SSC, 1 % SDS, 5 x Denharts, 10 % dextran sulfate, 100 μ g/ml salmon sperm DNA, and 32 p labeled probe and wash at 68 °C with 3, 2, 1, 0.5 or 0.1 x SSC and 0.1 % SDS.

30 According to another preferred embodiment of the present invention the polynucleotide or a portion thereof is at least 60 %, preferably at least 65 %, more preferably at least 70 %, more preferably at least 75 %, more preferably at least 80 %, more preferably at least 85 %, more preferably at least 90 %, most preferably, 95-100 % identical with SEQ ID NOs: 9, 13, 42, 43 or portions thereof as determined using the Bestfit procedure of the 35 DNA sequence analysis software package developed by the Genetic Computer Group (GCG) at the university of Wisconsin (gap creation penalty - 12, gap extension penalty - 4 - which are the default parameters).

According to another preferred embodiment of the present invention the polypeptide encoded by the polynucleotide sequence is as set forth in SEQ ID NOS:10, 14, 44 or portions thereof having heparanase catalytic activity. Such portions are expected to include amino acids Asp-Glu 224-5 225 (SEQ ID NO:10), which can serve as proton donors and glutamic acid 343 or 396 which can serve as a nucleophile.

According to another preferred embodiment of the present invention the polypeptide encoded by the polynucleotide sequence is at least 60 %, preferably at least 65 %, more preferably at least 70 %, more preferably at 10 75 %, more preferably at least 80 %, more preferably at least 85 %, more preferably at least 90 %, most preferably, 95-100 % homologous (both similar and identical acids) to SEQ ID NOS:10, 14, 44 or portions thereof as determined with the Smith-Waterman algorithm, using the Bioaccelerator platform developed by Compugene (gapop: 10.0, gapext: 0.5, matrix: 15 blosum62, see also the description to Figure 17).

Further according to the present invention there is provided a nucleic acid construct comprising the isolated nucleic acid described herein. The construct may and preferably further include an origin of replication and

20 The construct or vector can be of any type. It may be a phage which infects bacteria or a virus which infects eukaryotic cells. It may also be a plasmid, phagemid, cosmid, bacmid or an artificial chromosome.

Further according to the present invention there is provided a host 25 cell comprising the nucleic acid construct described herein. The host cell can be of any type. It may be a prokaryotic cell, an eukaryotic cell, a cell line, or a cell as a portion of an organism. The polynucleotide encoding heparanase can be permanently or transiently present in the cell. In other words, genetically modified cells obtained following stable or transient transfection, transformation or transduction are all within the scope of the 30 present invention. The polynucleotide can be present in the cell in low copy (say 1-5 copies) or high copy number (say 5-50 copies or more). It may be integrated in one or more chromosomes at any location or be present as an extrachromosomal material.

The present invention is further directed at providing a heparanase 35 over-expression system which includes a cell overexpressing heparanase catalytic activity. The cell may be a genetically modified host cell transiently or stably transfected or transformed with any suitable vector which includes a polynucleotide sequence encoding a polypeptide having

heparanase activity and a suitable promoter and enhancer sequences to direct over-expression of heparanase. However, the overexpressing cell may also be a product of an insertion (e.g., via homologous recombination) of a promoter and/or enhancer sequence downstream to the endogenous 5 heparanase gene of the expressing cell, which will direct over-expression from the endogenous gene.

The term "over-expression" as used herein in the specification and claims below refers to a level of expression which is higher than a basal level of expression typically characterizing a given cell under otherwise 10 identical conditions.

According to another aspect the present invention provides an antisense oligonucleotide comprising a polynucleotide or a polynucleotide analog of at least 10, preferably 11-15, more preferably 16-17, more 15 preferably 18, more preferably 19-25, more preferably 26-35, most preferably 35-100 bases being hybridizable *in vivo*, under physiological conditions, with a portion of a polynucleotide strand encoding a polypeptide having heparanase catalytic activity. The antisense oligonucleotide can be used for downregulating heparanase activity by *in vivo* administration thereof to a patient. As such, the antisense oligonucleotide according to the 20 present invention can be used to treat types of cancers which are characterized by impaired (over) expression of heparanase, and are dependent on the expression of heparanase for proliferating or forming metastases.

The antisense oligonucleotide can be DNA or RNA or even include 25 nucleotide analogs, examples of which are provided in the Background section hereinabove. The antisense oligonucleotide according to the present invention can be synthetic and is preferably prepared by solid phase synthesis. In addition, it can be of any desired length which still provides specific base pairing (e.g., 8 or 10, preferably more, nucleotides long) and it 30 can include mismatches that do not hamper base pairing under physiological conditions.

Further according to the present invention there is provided a pharmaceutical composition comprising the antisense oligonucleotide herein described and a pharmaceutically acceptable carrier. The carrier can be, for example, a liposome loadable with the antisense oligonucleotide. 35

According to a preferred embodiment of the present invention the antisense oligonucleotide further includes a ribozyme sequence. The ribozyme sequence serves to cleave a heparanase RNA molecule to which

the antisense oligonucleotide binds, to thereby downregulate heparanase expression.

Further according to the present invention there is provided an antisense nucleic acid construct comprising a promoter sequence and a 5 polynucleotide sequence directing the synthesis of an antisense RNA sequence of at least 10 bases being hybridizable *in vivo*, under physiological conditions, with a portion of a polynucleotide strand encoding a polypeptide having heparanase catalytic activity. Like the antisense oligonucleotide, the antisense construct can be used for downregulating heparanase activity by 10 *in vivo* administration thereof to a patient. As such, the antisense construct, like the antisense oligonucleotide, according to the present invention can be used to treat types of cancers which are characterized by impaired (over) expression of heparanase, and are dependent on the expression of heparanase for proliferating or forming metastases.

15 Thus, further according to the present invention there is provided a pharmaceutical composition comprising the antisense construct herein described and a pharmaceutically acceptable carrier. The carrier can be, for example, a liposome loadable with the antisense construct.

20 Formulations for topical administration may include, but are not limited to, lotions, ointments, gels, creams, suppositories, drops, liquids, sprays and powders. Conventional pharmaceutical carriers, aqueous, powder or oily bases, thickeners and the like may be necessary or desirable. Coated condoms, stents, active pads, and other medical devices may also be 25 useful. Compositions for oral administration include powders or granules, suspensions or solutions in water or non-aqueous media, sachets, capsules or tablets. Thickeners, diluents, flavorings, dispersing aids, emulsifiers or binders may be desirable. Formulations for parenteral administration may include, but are not limited to, sterile aqueous solutions which may also contain buffers, diluents and other suitable additives.

30 Dosing is dependent on severity and responsiveness of the condition to be treated, but will normally be one or more doses per day, week or month with course of treatment lasting from several days to several months or until a cure is effected or a diminution of disease state is achieved. Persons ordinarily skilled in the art can easily determine optimum dosages, 35 dosing methodologies and repetition rates.

Further according to the present invention there is provided a nucleic acid construct comprising a polynucleotide sequence functioning as a promoter, the polynucleotide sequence is derived from SEQ ID NO:42 and

includes at least nucleotides 2135-2635, preferably 2235-2635, more preferably 2335-2635, more preferably 2435-2635, most preferably 2535-2635 thereof, or SEQ ID NO:43 and includes at least nucleotides 1-420, preferably 120-420, more preferably 220-420, most preferably 320-420, 5 thereof. These nucleotides are shown in the example section that follows to direct the synthesis of a reporter gene in transformed cells. Thus, further according to the present invention there is provided a method of expressing a polynucleotide sequence comprising the step of ligating the polynucleotide sequence downstream to either of the promoter sequences 10 described herein. Heparanase promoters can be isolated from a variety of mammalian an other species by cloning genomic regions present 5' to the coding sequence thereof. This can be readily achievable by one ordinarily skilled in the art using the heparanase polynucleotides described herein, which are shown in the Examples section that follows to participate in 15 efficient cross species hybridization.

Further according to the present invention there is provided a recombinant protein comprising a polypeptide having heparanase catalytic activity. The protein according to the present invention include modifications known as post translational modifications, including, but not 20 limited to, proteolysis (e.g., removal of a signal peptide and of a pro- or preprotein sequence), methionine modification, glycosylation, alkylation (e.g., methylation), acetylation, etc. According to preferred embodiments the polypeptide includes at least a portion of SEQ ID NOs:10, 14 or 44, the portion has heparanase catalytic activity. According to preferred 25 embodiments of the present invention the protein is encoded by any of the above described isolated nucleic acids. Further according to the present invention there is provided a pharmaceutical composition comprising, as an active ingredient, the recombinant protein described herein.

The recombinant protein may be purified by any conventional 30 protein purification procedure close to homogeneity and/or be mixed with additives. The recombinant protein may be manufactured using any of the genetically modified cells described above, which include any of the expression nucleic acid constructs described herein. The recombinant protein may be in any form. It may be in a crystallized form, a dehydrated 35 powder form or in solution. The recombinant protein may be useful in obtaining pure heparanase, which in turn may be useful in eliciting anti-heparanase antibodies, either poly or monoclonal antibodies, and as a

screening active ingredient in an anti-heparanase inhibitors or drugs screening assay or system.

Further according to the present invention there is provided a method of identifying a chromosome region harboring a human heparanase gene in a chromosome spread. the method is executed implementing the following method steps, in which in a first step the chromosome spread (either interphase or metaphase spread) is hybridized with a tagged polynucleotide probe encoding heparanase. The tag is preferably a fluorescent tag. In a second step according to the method the chromosome spread is washed, thereby excess of non-hybridized probe is removed. Finally, signals associated with the hybridized tagged polynucleotide probe are searched for, wherein detected signals being indicative of a chromosome region harboring the human heparanase gene. One ordinarily skilled in the art would know how to use the sequences disclosed herein in suitable labeling reactions and how to use the tagged probes to detect, using *in situ* hybridization, a chromosome region harboring a human heparanase gene.

Further according to the present invention there is provided a method of *in vivo* eliciting anti-heparanase antibodies comprising the steps of administering a nucleic acid construct including a polynucleotide segment corresponding to at least a portion of SEQ ID NOs:9, 13 or 43 and a promoter for directing the expression of said polynucleotide segment *in vivo*. Accordingly, there is provided also a DNA vaccine for *in vivo* eliciting anti-heparanase antibodies comprising a nucleic acid construct including a polynucleotide segment corresponding to at least a portion of SEQ ID NOs:9, 13 or 43 and a promoter for directing the expression of said polynucleotide segment *in vivo*. The vaccine optionally further includes a pharmaceutically acceptable carrier, such as a virus, liposome or an antigen presenting cell. Alternatively, the vaccine is employed as a naked DNA vaccine

The present invention can be used to develop treatments for various diseases, to develop diagnostic assays for these diseases and to provide new tools for basic research especially in the fields of medicine and biology.

Specifically, the present invention can be used to develop new drugs to inhibit tumor cell metastasis, inflammation and autoimmunity. The identification of the *hpa* gene encoding for the heparanase enzyme enables the production of a recombinant enzyme in heterologous expression systems.

Furthermore, the present invention can be used to modulate bioavailability of heparin-binding growth factors, cellular responses to heparin-binding growth factors (e.g., bFGF, VEGF) and cytokines (e.g., IL-8), cell interaction with plasma lipoproteins, cellular susceptibility to viral, 5 protozoa and some bacterial infections, and disintegration of neurodegenerative plaques. Recombinant heparanase offers a potential treatment for wound healing, angiogenesis, restenosis, atherosclerosis, inflammation, neurodegenerative diseases (such as, for example, Genstmann-Straussler Syndrome, Creutzfeldt-Jakob disease, Scrapie and 10 Alzheimer's disease) and certain viral and some bacterial and protozoa infections. Recombinant heparanase can be used to neutralize plasma heparin, as a potential replacement of protamine.

As used herein, the term "modulate" includes substantially inhibiting, slowing or reversing the progression of a disease, substantially ameliorating 15 clinical symptoms of a disease or condition, or substantially preventing the appearance of clinical symptoms of a disease or condition. A "modulator" therefore includes an agent which may modulate a disease or condition. Modulation of viral, protozoa and bacterial infections includes any effect which substantially interrupts, prevents or reduces any viral, bacterial or 20 protozoa activity and/or stage of the virus, bacterium or protozoon life cycle, or which reduces or prevents infection by the virus, bacterium or protozoon in a subject, such as a human or lower animal.

As used herein, the term "wound" includes any injury to any portion 25 of the body of a subject including, but not limited to, acute conditions such as thermal burns, chemical burns, radiation burns, burns caused by excess exposure to ultraviolet radiation such as sunburn, damage to bodily tissues such as the perineum as a result of labor and childbirth, including injuries sustained during medical procedures such as episiotomies, trauma-induced injuries including cuts, those injuries sustained in automobile and other 30 mechanical accidents, and those caused by bullets, knives and other weapons, and post-surgical injuries, as well as chronic conditions such as pressure sores, bedsores, conditions related to diabetes and poor circulation, and all types of acne, etc.

Anti-heparanase antibodies, raised against the recombinant enzyme, 35 would be useful for immunodetection and diagnosis of micrometastases, autoimmune lesions and renal failure in biopsy specimens, plasma samples, and body fluids. Such antibodies may also serve as neutralizing agents for heparanase activity.

The genomic heparanase sequences described herein can be used to construct knock-in and knock-out constructs. Such constructs include a fragment of 10-20 Kb of a heparanase locus and a negative and a positive selection markers and can be used to provide heparanase knock-in and knock-out animal models by methods known to the skilled artisan. Such animal models can be used for studying the function of heparanase in developmental processes, and in normal as well as pathological processes. They can also serve as an experimental model for testing drugs and gene therapy protocols. The complementary heparanase sequence (cDNA) can be used to derive transgenic animals, overexpressing heparanase for same. Alternatively , if cloned in the antisense orientation, the complementary heparanase sequence (cDNA) can be used to derive transgenic animals under-expressing heparanase for same.

The heparanase promoter sequences described herein and other cis regulatory elements linked to the heparanase locus can be used to regulate the expression of genes. For example, these promoters can be used to direct the expression of a cytotoxic protein, such as TNF, in tumor cells. It will be appreciated that heparanase itself is abnormally expressed under the control of its own promoter and other cis acting elements in a variety of tumors, and its expression is correlated with metastasis. It is also abnormally highly expressed in inflammatory cells. The introns of the heparanase gene can be used for the same purpose, as it is known that introns, especially upstream introns include cis acting element which affect expression. A heparanase promoter fused to a reporter protein can be used to study/monitor its activity.

The polynucleotide sequences described herein can also be used to provide DNA vaccines which will elicit in vivo anti heparanase antibodies. Such vaccines can therefore be used to combat inflammatory and cancer.

Antisense oligonucleotides derived according to the heparanase sequences described herein, especially such oligonucleotides supplemented with ribozyme activity, can be used to modulate heparanase expression. Such oligonucleotides can be from the coding region, from the introns or promoter specific. Antisense heparanase nucleic acid constructs can similarly function, as well known in the art.

The heparanase sequences described herein can be used to study the catalytic mechanism of heparanase. Carefully selected site directed mutagenesis can be employed to provide modified heparanase proteins

having modified characteristics in terms of, for example, substrate specificity, sensitivity to inhibitors, etc.

While studying heparanase expression in a variety of cell types alternatively spliced transcripts were identified. Such transcripts if found 5 characteristic of certain pathological conditions can be used as markers for such conditions. Such transcripts are expected to direct the synthesis of heparanases with altered functions.

Additional objects, advantages, and novel features of the present 10 invention will become apparent to one ordinarily skilled in the art upon examination of the following examples, which are not intended to be limiting. Additionally, each of the various embodiments and aspects of the present invention as delineated hereinabove and as claimed in the claims section below finds experimental support in the following examples.

15

EXAMPLES

Generally, the nomenclature used herein and the laboratory 20 procedures in recombinant DNA technology described below are those well known and commonly employed in the art. Standard techniques are used for cloning, DNA and RNA isolation, amplification and purification. Generally 25 enzymatic reactions involving DNA ligase, DNA polymerase, restriction endonucleases and the like are performed according to the manufacturers' specifications. These techniques and various other techniques are generally performed according to Sambrook et al., *Molecular Cloning--A Laboratory Manual*, Cold Spring Harbor Laboratory, Cold Spring Harbor, N.Y. (1989), which is incorporated herein by reference. Other general references are provided throughout this document. The procedures therein are believed to be well known in the art and are provided for the convenience of the reader. All the information contained therein is incorporated herein by reference.

30

The following protocols and experimental details are referenced in the Examples that follow:

Purification and characterization of heparanase from a human hepatoma cell line and human placenta: A human hepatoma cell line (Sk-hep-1) was chosen as a source for purification of a human tumor-derived 35 heparanase. Purification was essentially as described in U.S. Pat. No. 5,362,641 to Fuks, which is incorporated by reference as if fully set forth

herein. Briefly, 500 liter, 5×10^{11} cells were grown in suspension and the heparanase enzyme was purified about 240,000 fold by applying the following steps: (i) cation exchange (CM-Sephadex) chromatography performed at pH 6.0, 0.3-1.4 M NaCl gradient; (ii) cation exchange (CM-Sephadex) chromatography performed at pH 7.4 in the presence of 0.1% CHAPS, 0.3-1.1 M NaCl gradient; (iii) heparin-Sepharose chromatography performed at pH 7.4 in the presence of 0.1% CHAPS, 0.35-1.1 M NaCl gradient; (iv) ConA-Sepharose chromatography performed at pH 6.0 in buffer containing 0.1 % CHAPS and 1 M NaCl, elution with 0.25 M α -methyl mannoside; and (v) HPLC cation exchange (Mono-S) chromatography performed at pH 7.4 in the presence of 0.1 % CHAPS, 0.25-1 M NaCl gradient.

Active fractions were pooled, precipitated with TCA and the precipitate subjected to SDS polyacrylamide gel electrophoresis and/or tryptic digestion and reverse phase HPLC. Tryptic peptides of the purified protein were separated by reverse phase HPLC (C8 column) and homogeneous peaks were subjected to amino acid sequence analysis.

The purified enzyme was applied to reverse phase HPLC and subjected to N-terminal amino acid sequencing using the amino acid sequencer (Applied Biosystems).

Cells: Cultures of bovine corneal endothelial cells (BCECs) were established from steer eyes as previously described (19, 38). Stock cultures were maintained in DMEM (1 g glucose/liter) supplemented with 10 % newborn calf serum and 5 % FCS. bFGF (1 ng/ml) was added every other day during the phase of active cell growth (13, 14).

Preparation of dishes coated with ECM: BCECs (second to fifth passage) were plated into 4-well plates at an initial density of 2×10^5 cells/ml, and cultured in sulfate-free Fisher medium plus 5 % dextran T-40 for 12 days. $\text{Na}_2^{35}\text{SO}_4$ (25 $\mu\text{Ci}/\text{ml}$) was added on day 1 and 5 after seeding and the cultures were incubated with the label without medium change. The subendothelial ECM was exposed by dissolving (5 min., room temperature) the cell layer with PBS containing 0.5 % Triton X-100 and 20 mM NH_4OH , followed by four washes with PBS. The ECM remained intact, free of cellular debris and firmly attached to the entire area of the tissue culture dish (19, 22).

To prepare soluble sulfate labeled proteoglycans (peak I material), the ECM was digested with trypsin (25 $\mu\text{g}/\text{ml}$, 6 h, 37 °C), the digest was concentrated by reverse dialysis and the concentrated material was applied

onto a Sepharose 6B gel filtration column. The resulting high molecular weight material ($Kav < 0.2$, peak I) was collected. More than 80 % of the labeled material was shown to be composed of heparan sulfate proteoglycans (11, 39).

5 **Heparanase activity:** Cells (1×10^6 /35-mm dish), cell lysates or conditioned media were incubated on top of 35 S-labeled ECM (18 h, 37 °C) in the presence of 20 mM phosphate buffer (pH 6.2). Cell lysates and conditioned media were also incubated with sulfate labeled peak I material (10-20 μ l). The incubation medium was collected, centrifuged (18,000 x g, 10 4 °C, 3 min.), and sulfate labeled material analyzed by gel filtration on a Sepharose CL-6B column (0.9 x 30 cm). Fractions (0.2 ml) were eluted with PBS at a flow rate of 5 ml/h and counted for radioactivity using Bio-fluor scintillation fluid. The excluded volume (V_0) was marked by blue dextran and the total included volume (V_t) by phenol red. The latter was 15 shown to comigrate with free sulfate (7, 11, 23). Degradation fragments of HS side chains were eluted from Sepharose 6B at $0.5 < Kav < 0.8$ (peak II) (7, 11, 23). A nearly intact HSPG released from ECM by trypsin - and, to a lower extent, during incubation with PBS alone - was eluted next to V_0 (10 $Kav < 0.2$, peak I). Recoveries of labeled material applied on the columns 20 ranged from 85 to 95 % in different experiments (11). Each experiment was performed at least three times and the variation of elution positions (Kav values) did not exceed +/- 15 %.

25 **Cloning of hpa cDNA:** cDNA clones 257548 and 260138 were obtained from the I.M.A.G.E Consortium (2130 Memorial Parkway SW, Huntsville, AL 35801). The cDNAs were originally cloned in *Eco*RI and *Not*I cloning sites in the plasmid vector pT3T7D-Pac. Although these clones are reported to be somewhat different, DNA sequencing demonstrated that these clones are identical to one another. Marathon RACE (rapid amplification of cDNA ends) human placenta (poly-A) cDNA 30 composite was a gift of Prof. Yossi Shiloh of Tel Aviv University. This composite is vector free, as it includes reverse transcribed cDNA fragments to which double, partially single stranded adapters are attached on both sides. The construction of the specific composite employed is described in reference 39a.

35 Amplification of hp3 PCR fragment was performed according to the protocol provided by Clontech laboratories. The template used for amplification was a sample taken from the above composite. The primers used for amplification were:

First step: 5'-primer: AP1: 5'-CCATCCTAATACGACTCACT ATAGGGC-3', SEQ ID NO:1; 3'-primer: HPL229: 5'-GTAGTGATGCCA TGTAACTGAATC-3', SEQ ID NO:2.

Second step: nested 5'-primer: AP2: 5'-ACTCACTATAGGGCTCG AGCGGC-3', SEQ ID NO:3; nested 3'- primer: HPL171: 5'-GCATCTTAGCCGTCTTCTTCG-3', SEQ ID NO:4. The HPL229 and HPL171 were selected according to the sequence of the EST clones. They include nucleotides 933-956 and 876-897 of SEQ ID NO:9, respectively.

PCR program was 94 °C - 4 min., followed by 30 cycles of 94 °C - 10 40 sec., 62 °C - 1 min., 72 °C - 2.5 min. Amplification was performed with Expand High Fidelity (Boehringer Mannheim). The resulting ca. 900 bp hp3 PCR product was digested with *Bfr*I and *Pvu*II. Clone 257548 (*phpa1*) was digested with *Eco*RI, followed by end filling and was then further digested with *Bfr*I. Thereafter the *Pvu*II - *Bfr*I fragment of the hp3 PCR 15 product was cloned into the blunt end - *Bfr*I end of clone *phpa1* which resulted in having the entire cDNA cloned in pT3T7-pac vector, designated *phpa2*.

RT-PCR: RNA was prepared using TRI-Reagent (Molecular research center Inc.) according to the manufacturer instructions. 1.25 µg 20 were taken for reverse transcription reaction using MuMLV Reverse transcriptase (Gibco BRL) and Oligo (dT)₁₅ primer, SEQ ID NO:5, (Promega). Amplification of the resultant first strand cDNA was performed with *Taq* polymerase (Promega). The following primers were used:

25 HPU-355: 5'-TTCGATCCCAAGAAGGAATCAAC-3', SEQ ID NO:6, nucleotides 372-394 in SEQ ID NOs:9 or 11.
HPL-229: 5'-GTAGTGATGCCATGTAACTGAATC-3', SEQ ID NO:7, nucleotides 933-956 in SEQ ID NOs:9 or 11.

PCR program: 94 °C - 4 min., followed by 30 cycles of 94 °C - 40 30 sec., 62 °C - 1 min., 72 °C - 1 min.

Alternatively, total RNA was prepared from cell cultures using Tri-reagent (Molecular Research Center, Inc.) according to the manufacturer recommendation. Poly A+ RNA was isolated from total RNA using mRNA separator (Clontech). Reverse transcription was performed with total RNA 35 using Superscript II (GibcoBRL). PCR was performed with Expand high fidelity (Boehringer Mannheim). Primers used for amplification were as follows:

Hpu-685, 5'-GAGCAGCCAGGTGAGGCCAAGAT-3', SEQ ID NO:24

Hpu-355, 5'-TTCGATCCCAAGAAGGAATCAAC-3', SEQ ID NO:25
Hpu 565, 5'-AGCTCTGTAGATGTGCTATACAC-3', SEQ ID NO:26
Hpl 967, 5'-TCAGATGCAAGCAGCAACTTGGC-3', SEQ ID NO:27
Hpl 171, 5'-GCATCTTAGCCGTCTTCTTCG-3', SEQ ID NO:28
5 Hpl 229, 5'-GTAGTGATGCCATGTAACTGAATC-3', SEQ ID NO:29

PCR reaction was performed as follows: 94 °C 3 minutes, followed by 32 cycles of 94 °C 40 seconds, 64 °C 1 minute, 72 °C 3 minutes, and one cycle 72 °C, 7 minutes.

10 **Expression of recombinant heparanase in insect cells:** Cells, High Five and Sf21 insect cell lines were maintained as monolayer cultures in SF900II-SFM medium (GibcoBRL).

15 **Recombinant Baculovirus:** Recombinant virus containing the *hpa* gene was constructed using the Bac to Bac system (GibcoBRL). The transfer vector pFastBac was digested with *Sal*I and *Not*I and ligated with a 1.7 kb fragment of *phpa2* digested with *Xho*I and *Not*I. The resulting plasmid was designated pFast*hpa2*. An identical plasmid designated pFast*hpa4* was prepared as a duplicate and both independently served for further experimentations. Recombinant bacmid was generated according to the instructions of the manufacturer with pFast*hpa2*, pFast*hpa4* and with pFastBac. The latter served as a negative control. Recombinant bacmid DNAs were transfected into Sf21 insect cells. Five days after transfection recombinant viruses were harvested and used to infect High Five insect cells, 3×10^6 cells in T-25 flasks. Cells were harvested 2 - 3 days after infection. 4×10^6 cells were centrifuged and resuspended in a reaction buffer containing 20 mM phosphate citrate buffer, 50 mM NaCl. Cells underwent three cycles of freeze and thaw and lysates were stored at -80 °C. Conditioned medium was stored at 4 °C.

20 **Partial purification of recombinant heparanase:** Partial purification of recombinant heparanase was performed by heparin-Sepharose column chromatography followed by Superdex 75 column gel filtration. Culture medium (150 ml) of Sf21 cells infected with pFhp4 virus was subjected to heparin-Sepharose chromatography. Elution of 1 ml fractions was performed with 0.35 - 2 M NaCl gradient in presence of 0.1 % CHAPS and 1 mM DTT in 10 mM sodium acetate buffer, pH 5.0. A 25 µl sample of each fraction was tested for heparanase activity. Heparanase activity was eluted at the range of 0.65 - 1.1 M NaCl (fractions 18-26, Figure 10a). 5 µl of each fraction was subjected to 15 % SDS-polyacrylamide gel electrophoresis followed by silver nitrate staining.

Active fractions eluted from heparin-Sepharose (Figure 10a) were pooled and concentrated (x 6) on YM3 cut-off membrane. 0.5 ml of the concentrated material was applied onto 30 ml Superdex 75 FPLC column equilibrated with 10 mM sodium acetate buffer, pH 5.0, containing 0.8 M NaCl, 1 mM DTT and 0.1 % CHAPS. Fractions (0.56 ml) were collected at a flow rate of 0.75 ml/min. Aliquots of each fraction were tested for heparanase activity and were subjected to SDS-polyacrylamide gel electrophoresis followed by silver nitrate staining (Figure 11b).

PCR amplification of genomic DNA: 94 °C 3 minutes, followed by 10 32 cycles of 94 °C 45 seconds, 64 °C 1 minute, 68 °C 5 minutes, and one cycle at 72 °C, 7 minutes. Primers used for amplification of genomic DNA included:

GHpu-L3 5'-AGGCACCCTAGAGATGTTCCAG-3', SEQ ID NO:30

GHpl-L6 5'-GAAGATTCTGTTCCATGACGTG-3', SEQ ID NO:31.

Screening of genomic libraries: A human genomic library in Lambda phage EMBLE3 SP6/T7 (Clontech, Paulo Alto, CA) was screened. 5 x 10⁵ plaques were plated at 5 x 10⁴ pfu/plate on NZCYM agar/top agarose plates. Phages were absorbed on nylon membranes in duplicates (Qiagen). Hybridization was performed at 65 °C in 5 x SSC, 5 x Denhart's, 20 10 % dextran sulfate, 100 µg/ml Salmon sperm, ³²P labeled probe (10⁶ cpm/ml). A 1.6 kb fragment, containing the entire *hpa* cDNA was labeled by random priming (Boehringer Mannheim). Following hybridization membranes were washed once with 2 x SSC, 0.1 % SDS at 65 °C for 20 minutes, and twice with 0.2 x SSC, 0.1 % SDS at 65 °C for 15 minutes. 25 Hybridizing plaques were picked, and plated at 100 pfu/plate. Hybridization was performed as above and single isolated positive plaques were picked.

Phage DNA was extracted using a Lambda DNA extraction kit (Qiagen). DNA was digested with *Xba*I and *Eco*RI, separated on 0.7 % 30 agarose gel and transferred to nylon membrane Hybond N+ (Amersham). Hybridization and washes were performed as above.

cDNA Sequence analysis: Sequence determinations were performed with vector specific and gene specific primers, using an automated DNA sequencer (Applied Biosystems, model 373A). Each nucleotide was read 35 from at least two independent primers.

Genomic sequence analysis: Large-scale sequencing was performed by Commonwealth Biotechnology Incorporation.

5 **Isolation of mouse hpa:** Mouse *hpa* cDNA was amplified from either Marathon ready cDNA library of mouse embryo or from mRNA isolated from mouse melanoma cell line BL6, using the Marathon RACE kit from Clontech. Both procedures were performed according to the manufacturer's recommendation.

Primers used for PCR amplification of mouse hpa:

Mhp1773 5'-CCACACTGAATGTAATACTGAAGTG-3', SEQ ID NO:32
Mhp1736 5'-CGAAGCTCTGGAACCTCGGCAAG-3', SEQ ID NO:33
Mhp183 5'-GCCAGCTGCAAAGGTGTTGGAC-3', SEQ ID NO:34
10 Mhp152 5'-AACACCTGCCTCATCACGACTTC-3', SEQ ID NO:35
Mhp1114 5'-GCCAGGCTGGCGTCGATGGTGA-3', SEQ ID NO:36
Mhp1103 5'-GTCGATGGTGATGGACAGGAAC-3', SEQ ID NO:37
Ap1 5'-GTAATACGACTCACTATAAGGGC-3', SEQ ID NO:38 -
(Genome walker)
15 Ap2 5'-ACTATAGGGCACGCGTGGT-3', SEQ ID NO:39 -
(Genome walker)
Ap1 5'-CCATCCTAATACGACTCACTATAAGGGC-3', SEQ ID NO:40 -
(Marathon RACE)
Ap2 5'-ACTCACTATAGGGCTCGAGCGGC-3', SEQ ID NO:41 -
20 (Marathon RACE)

25 **Southern analysis of genomic DNA:** Genomic DNA was extracted from animal or from human blood using Blood and cell culture DNA maxi kit (Qiagene). DNA was digested with *Eco*RI, separated by gel electrophoresis and transferred to a nylon membrane Hybond N+ (Amersham). Hybridization was performed at 68 °C in 6 x SSC, 1 % SDS, 5 x Denharts, 10 % dextran sulfate, 100 µg/ml salmon sperm DNA, and ³²P labeled probe. A 1.6 kb fragment, containing the entire *hpa* cDNA was used as a probe. Following hybridization, the membrane was washed with 3 x SSC, 0.1 % SDS, at 68 °C and exposed to X-ray film for 3 days.
30 Membranes were then washed with 1 x SSC, 0.1 % SDS, at 68 °C and were reexposed for 5 days.

35 **Construction of hpa promoter-GFP expression vector:** Lambda DNA of phage L3, was digested with *Sac*I and *Bgl*II, resulting in a 1712 bp fragment which contained the *hpa* promoter (877-2688 of SEQ ID NO:42). The pEGFP-1 plasmid (Clontech) was digested with *Bgl*II and *Sac*I and ligated with the 1712 bp fragment of the *hpa* promoter sequence. The resulting plasmid was designated phpEGL. A second *hpa* promoter-GFP plasmid was constructed containing a shorter fragment of the *hpa* promoter

region: phpEGL was digested with *Hind*III, and the resulting 1095 bp fragment (nucleotides 1593-2688 of SEQ ID NO:42) was ligated with *Hind*III digested pEGFP-1. The resulting plasmid was designated phpEGS.

Computer analysis of sequences: Homology searches were performed using several computer servers, and various databases. Blast 2.0 service, at the NCBI server was used to screen the protein database swplus and DNA databases such as GenBank, EMBL, and the EST databases. Blast 2.0 search was performed using the basic search option of the NCBI server. Sequence analysis and alignments were done using the DNA sequence analysis software package developed by the Genetic Computer Group (GCG) at the university of Wisconsin. Alignments of two sequences were performed using Bestfit (gap creation penalty - 12, gap extension penalty - 4). Protein homology search was performed with the Smith-Waterman algorithm, using the Bioaccelerator platform developed by Compugene. The protein database swplus was searched using the following parameters: gapop: 10.0, gapext: 0.5, matrix: blosum62. Blocks homology was performed using the Blocks WWW server developed at Fred Hutchinson Cancer Research Center in Seattle, Washington, USA. Secondary structure prediction was performed using the PHD server – Profile network Prediction Heidelberg. Fold recognition (threading) was performed using the UCLA-DOE structure prediction server. The method used for prediction was gonnet+predss. Alignment of three sequences was performed using the pileup application (gap creation penalty - 5, gap extension penalty - 1). Promoter analysis was performed using TSSW and TSSG programs (BCM Search Launcher Human Genome Center, Baylor College of Medicine, Houston TX).

EXAMPLE 1

Cloning of human hpa cDNA

Purified fraction of heparanase isolated from human hepatoma cells (SK-hep-1) was subjected to tryptic digestion and microsequencing. EST (Expressed Sequence Tag) databases were screened for homology to the back translated DNA sequences corresponding to the obtained peptides. Two EST sequences (accession Nos. N41349 and N45367) contained a DNA sequence encoding the peptide YGPDVGQPR (SEQ ID NO:8). These two sequences were derived from clones 257548 and 260138 (I.M.A.G.E Consortium) prepared from 8 to 9 weeks placenta cDNA library (Soares). Both clones which were found to be identical contained an insert

of 1020 bp which included an open reading frame (ORF) of 973 bp followed by a 3' untranslated region of 27 bp and a Poly A tail. No translation start site (AUG) was identified at the 5' end of these clones.

5 Cloning of the missing 5' end was performed by PCR amplification of DNA from a placenta Marathon RACE cDNA composite. A 900 bp fragment (designated hp3), partially overlapping with the identified 3' encoding EST clones was obtained.

10 The joined cDNA fragment, 1721 bp long (SEQ ID NO:9), contained an open reading frame which encodes, as shown in Figure 1 and SEQ ID NO:11, a polypeptide of 543 amino acids (SEQ ID NO:10) with a calculated molecular weight of 61,192 daltons. The 3' end of the partial cDNA inserts contained in clones 257548 and 260138 started at nucleotide G⁷²¹ of SEQ ID NO:9 and Figure 1.

15 As further shown in Figure 1, there was a single sequence discrepancy between the EST clones and the PCR amplified sequence, which led to an amino acid substitution from Tyr²⁴⁶ in the EST to Phe²⁴⁶ in the amplified cDNA. The nucleotide sequence of the PCR amplified cDNA fragment was verified from two independent amplification products. The new gene was designated *hpa*.

20 As stated above, the 3' end of the partial cDNA inserts contained in EST clones 257548 and 260138 started at nucleotide 721 of *hpa* (SEQ ID NO:9). The ability of the *hpa* cDNA to form stable secondary structures, such as stem and loop structures involving nucleotide stretches in the vicinity of position 721 was investigated using computer modeling. It was 25 found that stable stem and loop structures are likely to be formed involving nucleotides 698-724 (SEQ ID NO:9). In addition, a high GC content, up to 70 %, characterizes the 5' end region of the *hpa* gene, as compared to about only 40 % in the 3' region. These findings may explain the immature termination and therefore lack of 5' ends in the EST clones.

30 To examine the ability of the *hpa* gene product to catalyze degradation of heparan sulfate in an *in vitro* assay the entire open reading frame was expressed in insect cells, using the Baculovirus expression system. Extracts of cells, infected with virus containing the *hpa* gene, demonstrated a high level of heparan sulfate degradation activity, while 35 cells infected with a similar construct containing no *hpa* gene had no such activity, nor did non-infected cells. These results are further demonstrated in the following Examples.

EXAMPLE 2***Degradation of soluble ECM-derived HSPG***

Monolayer cultures of High Five cells were infected (72 h, 28 °C) with recombinant Bacoluvirus containing the pFast*hpa* plasmid or with 5 control virus containing an insert free plasmid. The cells were harvested and lysed in heparanase reaction buffer by three cycles of freezing and thawing. The cell lysates were then incubated (18 h, 37 °C) with sulfate labeled, ECM-derived HSPG (peak I), followed by gel filtration analysis (Sephadex 6B) of the reaction mixture.

10 As shown in Figure 2, the substrate alone included almost entirely high molecular weight (Mr) material eluted next to V_0 (peak I, fractions 5-20, $K_{av} < 0.35$). A similar elution pattern was obtained when the HSPG substrate was incubated with lysates of cells that were infected with control virus. In contrast, incubation of the HSPG substrate with lysates of cells 15 infected with the *hpa* containing virus resulted in a complete conversion of the high Mr substrate into low Mr labeled degradation fragments (peak II, fractions 22-35, $0.5 < K_{av} < 0.75$).

20 Fragments eluted in peak II were shown to be degradation products of heparan sulfate, as they were (i) 5- to 6-fold smaller than intact heparan sulfate side chains (K_{av} approx. 0.33) released from ECM by treatment with either alkaline borohydride or papain; and (ii) resistant to further 25 digestion with papain or chondroitinase ABC, and susceptible to deamination by nitrous acid (6, 11). Similar results (not shown) were obtained with Sf21 cells. Again, heparanase activity was detected in cells infected with the *hpa* containing virus (pF*hpa*), but not with control virus (pF). This result was obtained with two independently generated 30 recombinant viruses. Lysates of control not infected High Five cells failed to degrade the HSPG substrate.

35 In subsequent experiments, the labeled HSPG substrate was incubated with medium conditioned by infected High Five or Sf21 cells.

As shown in Figures 3a-b, heparanase activity, reflected by the conversion of the high Mr peak I substrate into the low Mr peak II which represents HS degradation fragments, was found in the culture medium of 35 cells infected with the pF*hpa*2 or pF*hpa*4 viruses, but not with the control pF1 or pF2 viruses. No heparanase activity was detected in the culture medium of control non-infected High Five or Sf21 cells.

The medium of cells infected with the pF*hpa*4 virus was passed through a 50 kDa cut off membrane to obtain a crude estimation of the

molecular weight of the recombinant heparanase enzyme. As demonstrated in Figure 4, all the enzymatic activity was retained in the upper compartment and there was no activity in the flow through (<50 kDa) material. This result is consistent with the expected molecular weight of the 5 *hpa* gene product.

In order to further characterize the *hpa* product the inhibitory effect of heparin, a potent inhibitor of heparanase mediated HS degradation (40) was examined.

As demonstrated in Figures 5a-b, conversion of the peak I substrate 10 into peak II HS degradation fragments was completely abolished in the presence of heparin.

Altogether, these results indicate that the heparanase enzyme is expressed in an active form by insect cells infected with Baculovirus containing the newly identified human *hpa* gene.

15

EXAMPLE 3

Degradation of HSPG in intact ECM

Next, the ability of intact infected insect cells to degrade HS in intact, naturally produced ECM was investigated. For this purpose, High 20 Five or Sf21 cells were seeded on metabolically sulfate labeled ECM followed by infection (48 h, 28 °C) with either the pF*hpa*4 or control pF2 viruses. The pH of the medium was then adjusted to pH 6.2-6.4 and the cells further incubated with the labeled ECM for another 48 h at 28 °C or 24 h at 37 °C. Sulfate labeled material released into the incubation medium 25 was analyzed by gel filtration on Sepharose 6B.

As shown in Figures 6a-b and 7a-b, incubation of the ECM with cells infected with the control pF2 virus resulted in a constant release of labeled material that consisted almost entirely (>90%) of high Mr fragments (peak I) eluted with or next to V_0 . It was previously shown that a proteolytic 30 activity residing in the ECM itself and/or expressed by cells is responsible for release of the high Mr material (6). This nearly intact HSPG provides a soluble substrate for subsequent degradation by heparanase, as also indicated by the relatively large amount of peak I material accumulating when the heparanase enzyme is inhibited by heparin (6, 7, 12, Figure 9). 35 On the other hand, incubation of the labeled ECM with cells infected with the pF*hpa*4 virus resulted in release of 60-70% of the ECM-associated radioactivity in the form of low Mr sulfate-labeled fragments (peak II, 0.5 < K_{av} < 0.75), regardless of whether the infected cells were incubated with

the ECM at 28 °C or 37 °C. Control intact non-infected Sf21 or High Five cells failed to degrade the ECM HS side chains.

In subsequent experiments, as demonstrated in Figures 8a-b, High Five and Sf21 cells were infected (96 h, 28 °C) with pFhpa4 or control pF1 viruses and the culture medium incubated with sulfate-labeled ECM. Low Mr HS degradation fragments were released from the ECM only upon incubation with medium conditioned by pFhpa4 infected cells. As shown in Figure 9, production of these fragments was abolished in the presence of heparin. No heparanase activity was detected in the culture medium of control, non-infected cells. These results indicate that the heparanase enzyme expressed by cells infected with the pFhpa4 virus is capable of degrading HS when complexed to other macromolecular constituents (i.e. fibronectin, laminin, collagen) of a naturally produced intact ECM, in a manner similar to that reported for highly metastatic tumor cells or activated cells of the immune system (6, 7).

EXAMPLE 4

Purification of recombinant human heparanase

The recombinant heparanase was partially purified from medium of pFhpa4 infected Sf21 cells by Heparin-Sepharose chromatography (Figure 10a) followed by gel filtration of the pooled active fractions over an FPLC Superdex 75 column (Figure 11a). A ~ 63 kDa protein was observed, whose quantity, as was detected by silver stained SDS-polyacrylamide gel electrophoresis, correlated with heparanase activity in the relevant column fractions (Figures 10b and 11b, respectively). This protein was not detected in the culture medium of cells infected with the control pF1 virus and was subjected to a similar fractionation on heparin-Sepharose (not shown).

EXAMPLE 5

Expression of the human hpa cDNA in various cell types, organs and tissues

Referring now to Figures 12a-e, RT-PCR was applied to evaluate the expression of the *hpa* gene by various cell types and tissues. For this purpose, total RNA was reverse transcribed and amplified. The expected 585 bp long cDNA was clearly demonstrated in human kidney, placenta (8 and 11 weeks) and mole tissues, as well as in freshly isolated and short termed (1.5-48 h) cultured human placental cytotrophoblastic cells (Figure 12a), all known to express a high heparanase activity (41). The *hpa*

transcript was also expressed by normal human neutrophils (Figure 12b). In contrast, there was no detectable expression of the *hpa* mRNA in embryonic human muscle tissue, thymus, heart and adrenal (Figure 12b). The *hpa* gene was expressed by several, but not all, human bladder carcinoma cell lines (Figure 12c), SK hepatoma (SK-hep-1), ovarian carcinoma (OV 1063), breast carcinoma (435, 231), melanoma and megakaryocytic (DAMI, CHRF) human cell lines (Figures 12d-e).

The above described expression pattern of the *hpa* transcript was determined to be in a very good correlation with heparanase activity levels determined in various tissues and cell types (not shown).

EXAMPLE 6

Isolation of an extended 5' end of hpa cDNA from human SK-hep1 cell line

The 5' end of *hpa* cDNA was isolated from human SK-hep1 cell line by PCR amplification using the Marathon RACE (rapid amplification of cDNA ends) kit (Clontech). Total RNA was prepared from SK-hep1 cells using the TRI-Reagent (Molecular research center Inc.) according to the manufacturer instructions. Poly A+ RNA was isolated using the mRNA separator kit (Clonetech).

The Marathon RACE SK-hep1 cDNA composite was constructed according to the manufacturer recommendations. First round of amplification was performed using an adaptor specific primer AP1: 5'-CCATCCTAATACG ACTCACTATAGGGC-3', SEQ ID NO:1, and a *hpa* specific antisense primer hpl-629: 5'-CCCCAGGAGCAGCAGCATCAG-3', SEQ ID NO:17, corresponding to nucleotides 119-99 of SEQ ID NO:9. The resulting PCR product was subjected to a second round of amplification using an adaptor specific nested primer AP2: 5'-ACTCACTATAGGGCTCGAGCGGC-3', SEQ ID NO:3, and a *hpa* specific antisense nested primer hpl-666 5'-AGGCTTCGAGCGCAGCAGCAT-3', SEQ ID NO:18, corresponding to nucleotides 83-63 of SEQ ID NO:9. The PCR program was as follows: a hot start of 94 °C for 1 minute, followed by 30 cycles of 90 °C - 30 seconds, 68 °C - 4 minutes. The resulting 300 bp DNA fragment was extracted from an agarose gel and cloned into the vector pGEM-T Easy (Promega). The resulting recombinant plasmid was designated pHPSK1.

The nucleotide sequence of the pHPSK1 insert was determined and it was found to contain 62 nucleotides of the 5' end of the placenta *hpa* cDNA

(SEQ ID NO:9) and additional 178 nucleotides upstream, the first 178 nucleotides of SEQ ID NOs:13 and 15.

5 A single nucleotide discrepancy was identified between the SK-hep1 cDNA and the placenta cDNA. The "T" derivative at position 9 of the placenta cDNA (SEQ ID NO:9), is replaced by a "C" derivative at the corresponding position 187 of the SK-hep1 cDNA (SEQ ID NO:13).

10 The discrepancy is likely to be due to a mutation at the 5' end of the placenta cDNA clone as confirmed by sequence analysis of several additional cDNA clones isolated from placenta, which like the SK-hep1 cDNA contained C at position 9 of SEQ ID NO:9.

15 The 5' extended sequence of the SK-hep1 *hpa* cDNA was assembled with the sequence of the *hpa* cDNA isolated from human placenta (SEQ ID NO:9). The assembled sequence contained an open reading frame which encodes, as shown in SEQ ID NOs:14 and 15, a polypeptide of 592 amino acids with a calculated molecular weight of 66,407 daltons. The open reading frame is flanked by 93 bp 5' untranslated region (UTR).

EXAMPLE 7

Isolation of the upstream genomic region of the hpa gene

20 The upstream region of the *hpa* gene was isolated using the Genome Walker kit (Clontech) according to the manufacturer recommendations. The kit includes five human genomic DNA samples each digested with a different restriction endonuclease creating blunt ends: *Eco*RV, *Scal*I, *Dra*I, *Pvu*II and *Ssp*I.

25 The blunt ended DNA fragments are ligated to partially single stranded adaptors. The Genomic DNA samples were subjected to PCR amplification using the adaptor specific primer and a gene specific primer. Amplification was performed with Expand High Fidelity (Boehringer Mannheim).

30 A first round of amplification was performed using the ap1 primer: 5'-G TAATACGACTCACTATAGGGC-3', SEQ ID NO:19, and the *hpa* specific antisense primer hpl-666: 5'-AGGCTTCGAGCGCAGCAGCAT-3', SEQ ID NO:18, corresponding to nucleotides 83 – 63 of SEQ ID NO:9. The PCR program was as follows: a hot start of 94 °C - 3 minutes, followed
35 by 36 cycles of 94 °C - 40 seconds, 67 °C - 4 minutes.

The PCR products of the first amplification were diluted 1:50. One μ l of the diluted sample was used as a template for a second amplification using a nested adaptor specific primer ap2: 5'-

ACTATAGGGCACCGCGTGGT-3', SEQ ID NO:20, and a *hpa* specific antisense primer hpl-690, 5'-CTTGGGCTCACC TGGCTGCTC-3', SEQ ID NO:21, corresponding to nucleotides 62-42 of SEQ ID NO:9. The resulting amplification products were analyzed using agarose gel electrophoresis. 5 Five different PCR products were obtained from the five amplification reactions. A DNA fragment of approximately 750 bp which was obtained from the *Ssp*I digested DNA sample was gel extracted. The purified fragment was ligated into the plasmid vector pGEM-T Easy (Promega). The resulting recombinant plasmid was designated pGHP6905 and the 10 nucleotide sequence of the *hpa* insert was determined.

15 A partial sequence of 594 nucleotides is shown in SEQ ID NO:16. The last nucleotide in SEQ ID NO:13 corresponds to nucleotide 93 in SEQ ID:13. The DNA sequence in SEQ ID NO:16 contains the 5' region of the *hpa* cDNA and 501 nucleotides of the genomic upstream region which are predicted to contain the promoter region of the *hpa* gene.

EXAMPLE 8

Expression of the 592 amino acids HPA polypeptide in a human 293 cell line

20 The 592 amino acids open reading frame (SEQ ID NOs:13 and 15) was constructed by ligation of the 110 bp corresponding to the 5' end of the SK-hep1 *hpa* cDNA with the placenta cDNA. More specifically the Marathon RACE - PCR amplification product of the placenta *hpa* DNA was digested with *Sac*I and an approximately 1 kb fragment was ligated into a 25 *Sac*I-digested pGHP6905 plasmid. The resulting plasmid was digested with *Ear*I and *Aat*II. The *Ear*I sticky ends were blunted and an approximately 280 bp *Ear*I/blunt-*Aat*II fragment was isolated. This fragment was ligated with pFast*hpa* digested with *Eco*RI which was blunt ended using Klenow fragment and further digested with *Aat*II. The resulting plasmid contained a 30 1827 bp insert which includes an open reading frame of 1776 bp, 31 bp of 3' UTR and 21 bp of 5' UTR. This plasmid was designated pFast*Lhpa*.

35 A mammalian expression vector was constructed to drive the expression of the 592 amino acids heparanase polypeptide in human cells. The *hpa* cDNA was excised from pFast*Lhpa* with *Bss*HII and *Not*I. The resulting 1850 bp *Bss*HII-*Not*I fragment was ligated to a mammalian expression vector pSI (Promega) digested with *Mlu*I and *Not*I. The resulting recombinant plasmid, pSI*hpaMet2* was transfected into a human 293 embryonic kidney cell line.

Transient expression of the 592 amino-acids heparanase was examined by western blot analysis and the enzymatic activity was tested using the gel shift assay. Both these procedures are described in length in U.S. Pat. application No. 09/071,739, filed May 1, 1998, which is incorporated by reference as if fully set forth herein. Cells were harvested 3 days following transfection. Harvested cells were re-suspended in lysis buffer containing 150 mM NaCl, 50 mM Tris pH 7.5, 1% Triton X-100, 1 mM PMSF and protease inhibitor cocktail (Boehringer Mannheim). 40 µg protein extract samples were used for separation on a SDS-PAGE. Proteins were transferred onto a PVDF Hybond-P membrane (Amersham). The membrane was incubated with an affinity purified polyclonal anti heparanase antibody, as described in U.S. Pat. application No. 09/071,739. A major band of approximately 50 kDa was observed in the transfected cells as well as a minor band of approximately 65 kDa. A similar pattern was observed in extracts of cells transfected with the pShpa as demonstrated in U.S. Pat. application No. 09/071,739. These two bands probably represent two forms of the recombinant heparanase protein produced by the transfected cells. The 65 kDa protein probably represents a heparanase precursor, while the 50 kDa protein is suggested herein to be the processed or mature form.

The catalytic activity of the recombinant protein expressed in the pShpaMet2 transfected cells was tested by gel shift assay. Cell extracts of transfected and of mock transfected cells were incubated overnight with heparin (6 µg in each reaction) at 37 °C, in the presence of 20 mM phosphate citrate buffer pH 5.4, 1 mM CaCl₂, 1 mM DTT and 50 mM NaCl. Reaction mixtures were then separated on a 10 % polyacrylamide gel. The catalytic activity of the recombinant heparanase was clearly demonstrated by a faster migration of the heparin molecules incubated with the transfected cell extract as compared to the control. Faster migration indicates the disappearance of high molecular weight heparin molecules and the generation of low molecular weight degradation products.

EXAMPLE 9

Chromosomal localization of the hpa gene

Chromosomal mapping of the hpa gene was performed utilizing a panel of monochromosomal human/CHO and human/mouse somatic cell hybrids, obtained from the UK HGMP Resource Center (Cambridge, England).

40 ng of each of the somatic cell hybrid DNA samples were subjected to PCR amplification using the *hpa* primers: hpu565 5'-AGCTCTGTAGATGTGC TATACAC-3', SEQ ID NO:22, corresponding to nucleotides 564-586 of SEQ ID NO:9 and an antisense primer hpl171 5'-5 GCATCTTAGCCGTCTTCTTCG-3', SEQ ID NO:23, corresponding to nucleotides 897-876 of SEQ ID NO:9.

The PCR program was as follows: a hot start of 94 °C – 3 minutes, followed by 7 cycles of 94 °C – 45 seconds, 66 °C – 1 minute, 68 °C – 5 minutes, followed by 30 cycles of 94 °C – 45 seconds, 62 °C – 1 minute, 68 10 °C – 5 minutes, and a 10 minutes final extension at 72 °C.

The reactions were performed with Expand long PCR (Boehringer Mannheim). The resulting amplification products were analyzed using agarose gel electrophoresis. As demonstrated in Figure 14, a single band of approximately 2.8 Kb was obtained from chromosome 4, as well as from 15 the control human genomic DNA. A 2.8 kb amplification product is expected based on amplification of the genomic *hpa* clone (data not shown). No amplification products were obtained neither in the control DNA samples of hamster and mouse nor in somatic hybrids of other human chromosome.

20

EXAMPLE 10

Human genomic clone encoding heparanase

Five plaques were isolated following screening of a human genomic library and were designated L3-1, L5-1, L8-1, L10-1 and L6-1. The phage 25 DNAs were analyzed by Southern hybridization and by PCR with *hpa* specific and vector specific primers. Southern analysis was performed with three fragments of *hpa* cDNA: a *Pvu*II-*Bam*HI fragment (nucleotides 32-450, SEQ ID NO:9), a *Bam*HI-*Nde*I fragment (nucleotides 451-1102, SEQ ID NO:9) and an *Nde*I-*Xho*I fragment (nucleotides 1103-1721, SEQ ID NO:9).

Following Southern analysis, phages L3, L6, L8 were selected for further analysis. A scheme of the genomic region and the relative position of the three phage clones is depicted in Figure 15. A 2 kb DNA fragment containing the gap between phages L6 and L3 was PCR amplified from 35 human genomic DNA with two gene specific primers GHpuL3 and GHplL6. The PCR product was cloned into the plasmid vector pGEM-T-easy (Promega).

Large scale DNA sequencing of the three Lambda clones and the amplified fragment was performed with Lambda purified DNA by primer walking. A nucleotide sequence of 44,898 bp was analyzed (Figure 16, SEQ ID NO:42). Comparison of the genomic sequence with that of *hpa* cDNA revealed 12 exons separated by 11 introns (Figures 15 an 16). The genomic organization of the *hpa* gene is depicted in Figure 15 (top). The sequence include the coding region from the first ATG to the stop codon which spans 39,113 nucleotides, 2742 nucleotides upstream of the first ATG and 3043 nucleotides downstream of the stop codon. Splice site consensus sequences were identified at exon/intron junctions.

EXAMPLE 11

Alternative splicing

Several minor RT-PCR products were obtained from various cell types, following amplification with *hpa* specific primers. Each one found to contain a deletion of one or two exons. Some of these PCR products contain ORFs, which encode potential shorter proteins.

Table 1 below summarizes the alternative spliced products isolated from various cell lines.

Fragments of similar sizes were obtained following amplification with two cell lines, placenta and platelets.

	Cell type	Nucleotides deleted	Exons deleted	ORF
25	Platelets	1047-1267	8, 9	+
	Platelets	1154-1267	9	-
	Platelets	289-435, 562-735	2, 4	-
	Sk-hep1, platelets, Zr75	562-735	4	+
	Sk-hep1 (hepatoma)	561-904	4, 5	-
30	Zr75 (breast carcinoma)	96-203	1 (partial)	+

EXAMPLE 12

Mouse and rat hpa

EST databases were screened for sequences homologous to the *hpa* gene. Three mouse EST's were identified (accession No. Aa177901, from mouse spleen, Aa067997 from mouse skin, Aa47943 from mouse embryo), assembled into a 824 bp cDNA fragment which contains a partial open reading frame (lacking a 5' end) of 629 bp and a 3' untranslated region of

195 bp (SEQ ID NO:12). As shown in Figure 13, the coding region is 80 % similar to the 3' end of the *hpa* cDNA sequence. These EST's are probably cDNA fragments of the mouse *hpa* homolog that encodes for the mouse heparanase.

5 Searching for consensus protein domains revealed an amino terminal homology between the heparanase and several precursor proteins such as Procollagen Alpha 1 precursor, Tyrosine-protein kinase-RYK, Fibulin-1, Insulin-like growth factor binding protein and several others. The amino terminus is highly hydrophobic and contains a potential trans-membrane 10 domain. The homology to known signal peptide sequences suggests that it could function as a signal peptide for protein localization.

The amino acid sequence of human heparanase was used to search for homologous sequences in the DNA and protein databases. Several 15 human EST's were identified, as well as mouse sequences highly homologous to human heparanase. The following mouse EST's were identified AA177901, AA674378, AA67997, AA047943, AA690179, AI122034, all sharing an identical sequence and correspond to amino acids 336-543 of the human heparanase sequence. The entire mouse heparanase cDNA was cloned, based on the nucleotide sequence of the mouse EST's. 20 PCR primers were designed and a Marathon RACE was performed using a Marathon cDNA library from 15 days mouse embryo (Clontech) and from BL6 mouse melanoma cell line. The mouse *hpa* homologous cDNA was isolated following several amplification steps. A 1.1 kb fragment was amplified from mouse embryo Marathon cDNA library. The first cycle of 25 amplification was performed with primers mhpl773 and Ap1 and the second cycle with primers mhpl736 and AP2. A 1.1 kb fragment was then amplified from BL6 Marathon cDNA library. The first cycle of amplification was performed with the primers mhpl152 and Ap1, and the second with mhpl83 and AP2. The combined sequence was homologous to 30 nucleotides 157 - 1702 of the human *hpa* cDNA, which encode amino acids 33-543. The 5' end of the mouse *hpa* gene was isolated from a mouse genomic DNA library using the Genome Walker kit (Clontech). An 0.9 kb fragment was amplified from a *Dra*I digested Genome walker DNA library. The first cycle of amplification was performed with primers mhpl114 and 35 Ap1 and the second with primers mhpl103 and AP2. The assembled sequence (SEQ ID NOs:43, 45) is 2396 nucleotides long. It contains an open reading frame of 1605 nucleotides, which encode a polypeptide of 535 amino acids (SEQ ID NOs:44, 45), 196 nucleotides of 3' untranslated

region (UTR), and an upstream sequence which includes the promoter region and the 5'-UTR of the mouse *hpa* cDNA.. According to two promoter predicting programs TSSW and TSSG, the transcription start site is localized to nucleotide 431 of SEQ ID NOs:43, 45, 163 nucleotides upstream of the first ATG codon. The 431 upstream genomic sequence contains the promoter region. A TATA box is predicted at position 394 of SEQ ID NOs:43, 45. The mouse and the human *hpa* genes share an average homology of 78 % between the nucleotide sequences and 81 % similarity between the deduced amino acid sequences.

Search for *hpa* homologous sequences, using the Blast 2.0 server revealed two EST's from rat: AI060284 (385 nucleotides, SEQ ID NO:46) which is homologous to the amino terminus (68 % similarity to amino acids 12-136) of human heparanase and AI237828 (541 nucleotides, SEQ ID NO:47) which is homologous to the carboxyl terminus (81 % similarity to amino acids 500-543) of human heparanase, and contains a 3'-UTR. A comparison between the human heparanase and the mouse and rat homologous sequences is demonstrated in Figure 17.

EXAMPLE 13

Prediction of heparanase active site

Homology search of heparanase amino acid sequence against the DNA and the protein databases revealed no significant homologies. The protein secondary structure as predicted by the PHD program consists of alternating alpha helices and beta sheets. The fold recognition server of UCLA predicted alpha/beta barrel structure, with under-threshold confidence.

Five of 15 proteins, which were predicted to have most similar folds, were glycosyl hydrolases from various organisms: 1xyza – xylanase from *Clostridium Thermocellum*, 1pbga – 6-phospho-beta- δ -galactosidase from *Lactococcus Lactis*, 1amy – alpha-amylase from Barley, 1ecea – endocellulase from *Acidothermus Cellulolyticus* and 1qbc – hexosaminidase alpha chain, glycosyl hydrolase.

Protein homology search using the bioaccelerator pulled out several proteins, including glycosyl hydrolyses such as beta-fructofuranosidase from *Vicia faba* (broad bean) and from potato, lactase phlorizin hydrolase from human, xylanases from *Clostridium thermocellum* and from *Streptomyces halstedii* and cellulase from *Clostridium thermocellum*. Blocks 9.3 database pulled out the active site of glycosyl hydrolases family

five, which includes cellulases from various bacteria and fungi. Similar active site motif is shared by several lysosomal acid hydrolases (63) and other glycosyl hydrolases. The common mechanism shared by these enzymes involves two glutamic acid residues, a proton donor and a nucleophile.

Despite the lack of an overall homology between the heparanase and other glycosyl hydrolases, the amino acid couple Asp-Glu (NE), which is characteristic of the proton donor of glycosyl hydrolyses of the GH-A clan, was found at positions 224-225 of the human heparanase protein sequence.

10 As in other clan members, this NE couple is located at the end of a β sheet.

Considering the relative location of the proton donor and the predicted secondary structure, the glutamic acid that functions as nucleophile is most likely located at position 343, or at position 396. Identification of the active site and the amino acids directly involved in hydrolysis opens the way for expression of the defined catalytic domain. In 15 addition, it will provide the tools for rational design of enzyme activity either by modification of the microenvironment or catalytic site itself.

EXAMPLE 14

20 *Expression of hpa antisense in mammalian cell lines*

A mammalian expression vector Hpa2Kepcdna3 was constructed in order to express *hpa* antisense in mammalian cells. *hpa* cDNA (1.7 kb *Eco*RI fragment) was cloned into the plasmid pCDNA3 in 3'>5' (antisense) orientation. The construct was used to transfet MBT2-T50 and T24P cell 25 lines. 2×10^5 cells in 35 mm plates were transfected using the Fugene protocol (Boehringer Mannheim). 48 hours after transfection cells were trypsinized and seeded in six well plates. 24 hours later G418 was added to initiate selection. The number of colonies per 35 mm plate following 3 weeks:

30

	Antisense	No insert
T24P	15	60
MBT-T50	1	6

35 The lower number of colonies obtained after transfection with *hpa* antisense, as compared with the control plasmid suggests that the introduction of *hpa* antisense interfere with cell growth. This experiment demonstrates the use of complementary antisense *hpa* DNA sequence to

control heparanase expression in cells. This approach may be used to inhibit expression of heparanase *in vivo*, in, for example, cancer cells and in other pathological processes in which heparanase is involved.

5

EXAMPLE 15

Zoo blot

Hpa cDNA was used as a probe to detect homologous sequences in human DNA and in DNA of various animals. The autoradiogram of the Southern analysis is presented in Figure 18. Several bands were detected in 10 human DNA, which correlated with the accepted pattern according to the genomic *hpa* sequence. Several intense bands were detected in all mammals, while faint bands were detected in chicken. This correlates with the phylogenetic relation between human and the tested animals. The intense bands indicate that *hpa* is conserved among mammals as well as in 15 more genetically distant organisms. The multiple bands patterns suggest that in all animals, like in human, the *hpa* locus occupy large genomic region. Alternatively, the various bands could represent homologous sequences and suggest the existence of a gene family, which can be isolated based on their homology to the human *hpa* reported herein. This 20 conservation was actually found, between the isolated human *hpa* cDNA and the mouse homologue.

EXAMPLE 16

Characterization of the hpa promoter

25 The DNA sequence upstream of the *hpa* first ATG was subjected to computational analysis in order to localize the predicted transcription start site and to identify potential transcription factors binding sites. Recognition of human PolII promoter region and start of transcription were predicted using the TSSW and TSSG programs. Both programs identified a promoter 30 region upstream of the coding region. TSSW pointed at nucleotide 2644 and TSSG at 2635 of SEQ ID NO:42. These two predicted transcription start sites are located 4 and 13 nucleotides upstream of the longest *hpa* cDNA isolated by RACE.

35 A *hpa* promoter-GFP reporter vector was constructed in order to investigate the regulation of *hpa* transcription. Two constructs were made, containing 1.8 kb and 1.1 kb of the *hpa* promoter region. The reporter vector was transfected into T50-mouse bladder carcinoma cells. Cells transfected with both constructs exhibited green fluorescence, which

indicated the promoter activity of the genomic sequence upstream of the *hpa*-coding region. This reporter vector, enables the monitoring of *hpa* promoter activity, at various conditions and in different cell types and to characterize the factors involved regulation of *hpa* expression.

5

Although the invention has been described in conjunction with specific embodiments thereof, it is evident that many alternatives, modifications and variations will be apparent to those skilled in the art. Accordingly, it is intended to embrace all such alternatives, modifications 10 and variations that fall within the spirit and broad scope of the appended claims.

LIST OF REFERENCES

1. Wight, T.N., Kinsella, M.G., and Qwarnstrom, E.E. (1992). The role of proteoglycans in cell adhesion, migration and proliferation. *Curr. Opin. Cell Biol.*, 4, 793-801.
2. Jackson, R.L., Busch, S.J., and Cardin, A.L. (1991). Glycosaminoglycans: Molecular properties, protein interactions and role in physiological processes. *Physiol. Rev.*, 71, 481-539.
3. Wight, T.N. (1989). Cell biology of arterial proteoglycans. *Arteriosclerosis*, 9, 1-20.
4. Kjellen, L., and Lindahl, U. (1991). Proteoglycans: structures and interactions. *Annu. Rev. Biochem.*, 60, 443-475.
5. Ruoslahti, E., and Yamaguchi, Y. (1991). Proteoglycans as modulators of growth factor activities. *Cell*, 64, 867-869.
6. Vlodavsky, I., Eldor, A., Haimovitz-Friedman, A., Matzner, Y., Ishai-Michaeli, R., Levi, E., Bashkin, P., Lider, O., Naparstek, Y., Cohen, I.R., and Fuks, Z. (1992). Expression of heparanase by platelets and circulating cells of the immune system: Possible involvement in diapedesis and extravasation. *Invasion & Metastasis*, 12, 112-127.
7. Vlodavsky, I., Mohsen, M., Lider, O., Ishai-Michaeli, R., Ekre, H.-P., Svahn, C.M., Vigoda, M., and Peretz, T. (1995). Inhibition of tumor metastasis by heparanase inhibiting species of heparin. *Invasion & Metastasis*, 14, 290-302.
8. Nakajima, M., Irimura, T., and Nicolson, G.L. (1988). Heparanase and tumor metastasis. *J. Cell. Biochem.*, 36, 157-167.
9. Nicolson, G.L. (1988). Organ specificity of tumor metastasis: Role of preferential adhesion, invasion and growth of malignant cells at specific secondary sites. *Cancer Met. Rev.*, 7, 143-188.

10. Liotta, L.A., Rao, C.N., and Barsky, S.H. (1983). Tumor invasion and the extracellular matrix. *Lab. Invest.*, 49, 639-649.
11. Vlodavsky, I., Fuks, Z., Bar-Ner, M., Ariav, Y., and Schirrmacher, V. (1983). Lymphoma cell mediated degradation of sulfated proteoglycans in the subendothelial extracellular matrix: Relationship to tumor cell metastasis. *Cancer Res.*, 43, 2704-2711.
12. Vlodavsky, I., Ishai-Michaeli, R., Bar-Ner, M., Fridman, R., Horowitz, A.T., Fuks, Z. and Biran, S. (1988). Involvement of heparanase in tumor metastasis and angiogenesis. *Isr. J. Med.*, 24, 464-470.
13. Vlodavsky, I., Liu, G.M., and Gospodarowicz, D. (1980). Morphological appearance, growth behavior and migratory activity of human tumor cells maintained on extracellular matrix vs. plastic. *Cell*, 19, 607-616.
14. Gospodarowicz, D., Delgado, D., and Vlodavsky, I. (1980). Permissive effect of the extracellular matrix on cell proliferation in-vitro. *Proc. Natl. Acad. Sci. USA.*, 77, 4094-4098.
15. Bashkin, P., Doctrow, S., Klagsbrun, M., Svahn, C.M., Folkman, J., and Vlodavsky, I. (1989). Basic fibroblast growth factor binds to subendothelial extracellular matrix and is released by heparitinase and heparin-like molecules. *Biochemistry*, 28, 1737-1743.
16. Parish, C.R., Coombe, D.R., Jakobsen, K.B., and Underwood, P.A. (1987). Evidence that sulphated polysaccharides inhibit tumor metastasis by blocking tumor cell-derived heparanase. *Int. J. Cancer*, 40, 511-517.
- 16a. Vlodavsky, I., Hua-Quan Miao., Benezra, M., Lider, O., Bar-Shavit, R., Schmidt, A., and Peretz, T. (1997). Involvement of the extracellular matrix, heparan sulfate proteoglycans and heparan sulfate degrading enzymes in angiogenesis and metastasis. In: *Tumor Angiogenesis*. Eds. C.E. Lewis, R. Bicknell & N. Ferrara. Oxford University Press, Oxford UK, pp. 125-140.

17. Burgess, W.H., and Maciag, T. (1989). The heparin-binding (fibroblast) growth factor family of proteins. *Annu. Rev. Biochem.*, 58, 575-606.
18. Folkman, J., and Klagsbrun, M. (1987). Angiogenic factors. *Science*, 235, 442-447.
19. Vlodavsky, I., Folkman, J., Sullivan, R., Fridman, R., Ishai-Michaeli, R., Sasse, J., and Klagsbrun, M. (1987). Endothelial cell-derived basic fibroblast growth factor: Synthesis and deposition into subendothelial extracellular matrix. *Proc. Natl. Acad. Sci. USA*, 84, 2292-2296.
20. Folkman, J., Klagsbrun, M., Sasse, J., Wadzinski, M., Ingber, D., and Vlodavsky, I. (1980). A heparin-binding angiogenic protein - basic fibroblast growth factor - is stored within basement membrane. *Am. J. Pathol.*, 130, 393-400.
21. Cardon-Cardo, C., Vlodavsky, I., Haimovitz-Friedman, A., Hicklin, D., and Fuks, Z. (1990). Expression of basic fibroblast growth factor in normal human tissues. *Lab. Invest.*, 63, 832-840.
22. Ishai-Michaeli, R., Svahn, C.-M., Chajek-Shaul, T., Korner, G., Ekre, H.-P., and Vlodavsky, I. (1992). Importance of size and sulfation of heparin in release of basic fibroblast factor from the vascular endothelium and extracellular matrix. *Biochemistry*, 31, 2080-2088.
23. Ishai-Michaeli, R., Eldor, A., and Vlodavsky, I. (1990). Heparanase activity expressed by platelets, neutrophils and lymphoma cells releases active fibroblast growth factor from extracellular matrix. *Cell Reg.*, 1, 833-842.
24. Vlodavsky, I., Bar-Shavit, R., Ishai-Michaeli, R., Bashkin, P., and Fuks, Z. (1991). Extracellular sequestration and release of fibroblast growth factor: a regulatory mechanism? *Trends Biochem. Sci.*, 16, 268-271.
25. Vlodavsky, I., Bar-Shavit, R., Korner, G., and Fuks, Z. (1993). Extracellular matrix-bound growth factors, enzymes and plasma proteins. In *Basement membranes: Cellular and molecular aspects* (eds.

D.H. Rohrbach and R. Timpl), pp327-343. Academic press Inc., Orlando, Fl.

26. Yayon, A., Klagsbrun, M., Esko, J.D., Leder, P., and Ornitz, D.M. (1991). Cell surface, heparin-like molecules are required for binding of basic fibroblast growth factor to its high affinity receptor. *Cell*, 64, 841-848.

27. Spivak-Kroizman, T., Lemmon, M.A., Dikic, I., Ladbury, J.E., Pinchasi, D., Huang, J., Jaye, M., Crumley, G., Schlessinger, J., and Lax, I. (1994). Heparin-induced oligomerization of FGF molecules is responsible for FGF receptor dimerization, activation, and cell proliferation. *Cell*, 79, 1015-1024.

28. Ornitz, D.M., Herr, A.B., Nilsson, M., West, a., J., Svahn, C.-M., and Waksman, G. (1995). FGF binding and FGF receptor activation by synthetic heparan-derived di- and trisaccharides. *Science*, 268, 432-436.

29. Gitay-Goren, H., Soker, S., Vlodavsky, I., and Neufeld, G. (1992). Cell surface associated heparin-like molecules are required for the binding of vascular endothelial growth factor (VEGF) to its cell surface receptors. *J. Biol. Chem.*, 267, 6093-6098.

30. Lider, O., Baharav, E., Mekori, Y., Miller, T., Naparstek, Y., Vlodavsky, I., and Cohen, I.R. (1989). Suppression of experimental autoimmune diseases and prolongation of allograft survival by treatment of animals with heparinoid inhibitors of T lymphocyte heparanase. *J. Clin. Invest.*, 83, 752-756.

31. Lider, O., Cahalon, L., Gilat, D., Hershkovitz, R., Siegel, D., Margalit, R., Shoseyov, O., and Cohn, I.R. (1995). A disaccharide that inhibits tumor necrosis factor α is formed from the extracellular matrix by the enzyme heparanase. *Proc. Natl. Acad. Sci. USA*, 92, 5037-5041.

31a. Rapraeger, A., Krufka, A., and Olwin, B.R. (1991). Requirement of heparan sulfate for bFGF-mediated fibroblast growth and myoblast differentiation. *Science*, 252, 1705-1708.

32. Eisenberg, S., Sehayek, E., Olivecrona, T., and Vlodavsky, I. (1992). Lipoprotein lipase enhances binding of lipoproteins to heparan sulfate on cell surfaces and extracellular matrix. *J. Clin. Invest.*, 90, 2013-2021.

33. Shieh, M-T., Wundunn, D., Montgomery, R.I., Esko, J.D., and Spear, P.G. J. (1992). Cell surface receptors for herpes simplex virus are heparan sulfate proteoglycans. *J Cell Biol.*, 116, 1273-1281.

33a. Chen, Y., Maguire, T., Hileman, R.E., Fromm, J.R., Esko, J.D., Linhardt, R.J., and Marks, R.M. (1997). Dengue virus infectivity depends on envelope protein binding to target cell heparan sulfate. *Nature Medicine* 3, 866-871.

33b. Putnak, J.R., Kanessa-Thasan, N., and Innis, B.L. (1997). A putative cellular receptor for dengue viruses. *Nature Medicine* 3, 828-829.

34. Narindrasorasak, S., Lowery, D., Gonzalez-DeWhitt, P., Poorman, R.A., Greenberg, B., Kisilevsky, R. (1991). High affinity interactions between the Alzheimer's beta-amyloid precursor protein and the basement membrane form of the heparan sulfate proteoglycan. *J. Biol. Chem.*, 266, 12878-83.

35. Ross, R. (1993). The pathogenesis of atherosclerosis: a perspective for the 1990s. *Nature (Lond.)*, 362:801-809.

36. Zhong-Sheng, J., Walter, J., Brecht, R., Miranda, D., Mahmood Hussain, M., Innerarity, T.L. and Mahley, W.R. (1993). Role of heparan sulfate proteoglycans in the binding and uptake of apolipoprotein E-enriched remnant lipoproteins by cultured cells. *J. Biol. Chem.*, 268, 10160-10167.

37. Ernst, S., Langer, R., Cooney, Ch.L., and Sasisekharan, R. (1995). Enzymatic degradation of glycosaminoglycans. Critical Reviews in Biochemistry and Molecular Biology, 30(5), 387-444.

38. Gospodarowicz, D., Mescher, AL., Birdwell, CR. (1977). Stimulation of corneal endothelial cell proliferation in vitro by fibroblast and epidermal growth factors. *Exp Eye Res* 25, 75-89.

39. Haimovitz-Friedman, A., Falcone, D.J., Eldor, A., Schirrmacher, V., Vlodavsky, I., and Fuks, Z. (1991) Activation of platelet heparitinase by tumor cell-derived factors. *Blood*, 78, 789-796.

39a. Savitsky, K., Platzer, M., Uziel, T., Gilad, S., Sartiel, A., Rosenthal, A., Elroy-Stein, O., Siloh, Y. and Rotman, G. (1997). Ataxiatelangiectasia: structural diversity of untranslated sequences suggests complex post-translational regulation of ATM gene expression. *Nucleic Acids Res.* 25(9), 1678-1684.

40. Bar-Ner, M., Eldor, A., Wasserman, L., Matzner, Y., and Vlodavsky, I. (1987). Inhibition of heparanase mediated degradation of extracellular matrix heparan sulfate by modified and non-anticoagulant heparin species. *Blood*, 70, 551-557.

41. Goshen, R., Hochberg, A., Korner, G., Levi, E., Ishai-Michaeli, R., Elkin, M., de Grot, N., and Vlodavsky, I. (1996). Purification and characterization of placental heparanase and its expression by cultured cytotrophoblasts. *Mol. Human Reprod.*, 2, 679-684.

42. Korb M., Ke Y. and Johnson L.F. (1993) Stimulation of gene expression by introns: conversion of an inhibitory intron to a stimulatory intron by alteration of the splice donor sequence. *Nucleic Acids Res.*, 25;21(25):5901-8.

43. Zheng B., Qiu X.Y., Tan M., Xing Y.N., Lo D., Xue J.L. and Qiu X.F. (1997) Increment of hFIX expression with endogenous intron 1 in vitro. *Cell Res.*, 7(1):21-29.

44. Kurachi S., Hitomi Y., Furukawa M. and Kurachi K. (1995) Role of intron I in expression of the human factor IX gene. *J. Biol. Chem.* 10, 270(10):5276-5281.

45. Shekhar P.V. and Miller F.R. (1994-5) Correlation of differences in modulation of ras expression with metastatic competence of mouse mammary tumor subpopulations. *Invasion Metastasis*, 14(1-6):27-37.

46. Zhou G., Garofalo S., Mukhopadhyay K., Lefebvre V., Smith C.N., Eberspaecher H. and de Crombrugghe B. (1995) A 182 bp fragment of the mouse pro alpha 1(II) collagen gene is sufficient to direct chondrocyte expression in transgenic mice. *J. Cell Sci.*, 108 (Pt 12):3677-3684.

47. Hormuzdi S.G., Penttinen R., Jaenisch R. and Bornstein P. (1998) A gene-targeting approach identifies a function for the first intron in expression of the alpha1(I) collagen gene. *Mol. Cell*, 18(6):3368-3375.

48. Kang Y.K., Lee C.S., Chung A.S. and Lee K.K. (1998) Prolactin-inducible enhancer activity of the first intron of the bovine beta-casein gene. *Mol. Cells*, 30;8(3):259-265.

49. Chow Y.H., O'Brodovich H., Plumb J., Wen Y., Sohn K.J., Lu Z., Zhang F., Lukacs G.L., Tanswell A.K., Hui C.C., Buchwald M. and Hu J. (1997) Development of an epithelium-specific expression cassette with human DNA regulatory elements for transgene expression in lung airways. *Proc. Natl. Acad. Sci. USA*, 23;94(26):14695-14700.

50. Gottschalk U. and Chan S. (1998) Somatic gene therapy. Present situation and future perspective. *Arzneimittelforschung*, 48(11):1111-1120.

51. Ye S., Cole-Strauss A.C., Frank B. and Kmiec E.B. (1998) Targeted gene correction: a new strategy for molecular medicine. *Mol. Med. Today*, 4(10):431-437.

52. Lai L., and Lien Y. (1999) Homologous recombination based gene therapy. *Exp. Nephrol.*, 7(1):11-14.

53. Yazaki N., Fujita H., Ohta M., Kawasaki T. and Itoh N. (1993) The structure and expression of the FGF receptor-1 mRNA isoforms in rat tissues. *Biochim. Biophys. Acta.*, 20;1172(1-2):37-42.

54. Le Fur N., Kelsall S.R., Silvers W.K. and Mintz B. (1997) Selective increase in specific alternative splice variants of tyrosinase in murine melanomas: a projected basis for immunotherapy. *Proc. Natl. Acad. Sci. USA*, 13;94(10):5332-5337.

55. Miyake H., Okamoto I., Hara I., Gohji K., Yamanaka K., Arakawa S., Kamidono S. and Saya H. (1998) Highly specific and sensitive detection of malignancy in urine samples from patients with urothelial cancer by CD44v8-10/CD44v10 competitive RT-PCR. *Int. J. Cancer*, 18;79(6):560-564.

56. Guriec N., Marcellin L., Gairard B., Calderoli H., Wilk A., Renaud R., Bergerat J.P. and Oberling F. (1996) CD44 exon 6 expression as a possible early prognostic factor in primary node negative breast carcinoma. *Clin. Exp. Metastasis*, 14(5):434-439.

57. Gewirtz A.M., Sokol D.L. and Ratajczak M.Z. (1998) Nucleic acid therapeutics: state of the art and future prospects. *Blood*, 1;92(3):712-736.

58. Hida K., Shindoh M., Yasuda M., Hanzawa M., Funaoka K., Kohgo T., Amemiya A., Totsuka Y., Yoshida K. and Fujinaga K (1997) Antisense E1AF transfection restrains oral cancer invasion by reducing matrix metalloproteinase activities. *Am. J. Pathol.* 150(6):2125-2132.

59. Shastry B.S. (1998) Gene disruption in mice: models of development and disease. *Mol. Cell. Biochem.* 1998 Apr;181(1-2):163-179.

60. Carpentier A.F., Rosenfeld M.R., Delattre J.Y., Whalen R.G., Posner J.B. and Dalmau J. (1998) DNA vaccination with HuD inhibits growth of a neuroblastoma in mice. *Clin. Cancer Res.*, 4(11):2819-2824.

61. Lai W.C. and Bennett M. (1998) DNA vaccines. *Crit. Rev. Immunol.*, 18(5):449-484.

62. Welch P.J., Barber J.R., and Wong-Staal F. (1998) Expression of ribozymes in gene transfer systems to modulate target RNA levels. *Curr. Opin. Biotechnol.*, 9(5):486-496.

63. Durand P., Lehn P., Callebaunt I., Fabrega S., Henrissat B. and Mornon J.P. (1997) Active-site motifs of lysosomal acid hydrolyses: invariant features of clan GH-A glycosyl hydrolases deduced from hydrophobic cluster analysis. *Glycobiology*, 7(2):277-284.

64. Thuong and Helene (1993) Sequence specific recognition and modification of double helical DNA by oligonucleotides Angev. Chem. Int. Ed. Engl. 32:666

65. Dash P., Lotan I., Knapp M., Kandel E.R. and Goelet P. (1987) Selective elimination of mRNAs in vivo: complementary oligodeoxynucleotides promote RNA degradation by an RNase H-like activity. *Proc. Natl. Acad. Sci. USA*, 84:7896.

66. Chiang M.Y., Chan H., Zounes M.A., Freier S.M., Lima W.F. and Bennett C.F. (1991) Antisense oligonucleotides inhibit intercellular adhesion molecule 1 expression by two distinct mechanisms. *J. Biol. Chem.* 266:18162-71.

5 67. Paterson Paterson B.M, Roberts B.E and Kuff EL . (1977) Structural gene identification and mapping by DNA-mRNA hybrid-arrested cell-free translation. *Proc. Natl. Acad. Sci. USA*, 74:4370.

68. Cohen (1992) Oligonucleotide therapeutics. *Trends in Biotechnology*, 10:87.

69. Szczylak et al (1991) Selective inhibition of leukemia cell proliferation by BCR-ABL antisense oligodeoxynucleotides. *Science* 253:562.

70. Calabretta et al. (1991) Normal and leukemic hematopoietic cell manifest differential sensitivity to inhibitory effects of c-myc antisense

oligodeoxynucleotides: an in vitro study relevant to bone marrow purging. Proc. Natl. Acad. Sci. USA 88:2351.

71. Heikhila et al. (1987) A c-myc antisense oligodeoxynucleotide inhibits entry into S phase but not progress from G(0) to G(1). Nature, 328:445.

72. Reed et al. (1990) Antisense mediated inhibition of BCL2 prooncogene expression and leukemic cell growth and survival: comparison of phosphodiester and phosphorothioate oligodeoxynucleotides. Cancer Res. 50:6565.

73. Burch and Mahan (1991) Oligodeoxynucleotides antisense to the interleukin I receptor m RNA block the effects of interleukin I in cultured murine and human fibroblasts and in mice. J. Clin. Invest. 88:1190.

74. Agrawal (1992) Antisense oligonucleotides as antiviral agents. TIBTECH 10:152.

75. Uhlmann et al. (1990) Chem. Rev. 90:544.

76. Cook (1991) Medicinal chemistry of antisense oligonucleotides - future opportunities. Anti-Cancer Drug Design 6:585.

77. Biotechnology research news (1993) Can DNA mimics improve on the real thing? Science 262:1647.

WHAT IS CLAIMED IS:

1. An isolated nucleic acid comprising a genomic, complementary or composite polynucleotide sequence encoding a polypeptide having heparanase catalytic activity.
2. The isolated nucleic acid of claim 1, wherein said polynucleotide or a portion thereof is hybridizable with SEQ ID NOs: 9, 13, 42, 43 or a portion thereof at 68 °C in 6 x SSC, 1 % SDS, 5 x Denharts, 10 % dextran sulfate, 100 µg/ml salmon sperm DNA, and ^{32}p labeled probe and wash at 68 °C with 3 x SSC and 0.1 % SDS.
3. The isolated nucleic acid of claim 1, wherein said polynucleotide or a portion thereof is at least 60 % identical with SEQ ID NOs: 9, 13, 42, 43 or portions thereof as determined using the Bestfit procedure of the DNA sequence analysis software package developed by the Genetic Computer Group (GCG) at the university of Wisconsin (gap creation penalty - 12, gap extension penalty - 4).
4. The isolated nucleic acid of claim 1, wherein said polypeptide is as set forth in SEQ ID NOs:10, 14, 44 or portions thereof.
5. The isolated nucleic acid of claim 1, wherein said polypeptide is at least 60 % homologous to SEQ ID NOs:10, 14, 44 or portions thereof as determined with the Smith-Waterman algorithm, using the Bioaccelerator platform developed by Compugene (gapop: 10.0, gapext: 0.5, matrix: blosum62).
6. A nucleic acid construct comprising the isolated nucleic acid of claim 1.
7. A host cell comprising the nucleic acid construct of claim 6.
8. An antisense oligonucleotide comprising a polynucleotide or a polynucleotide analog of at least 10 bases being hybridizable *in vivo*, under physiological conditions, with a portion of a polynucleotide strand encoding a polypeptide having heparanase catalytic activity.

9. The antisense oligonucleotide of claim 8, wherein said polynucleotide strand encoding said polypeptide having heparanase catalytic activity is as set forth in SEQ ID NOs: 9, 13, 42, or 43.

10. The antisense oligonucleotide of claim 8, wherein said polypeptide having heparanase catalytic activity is as set forth in SEQ ID NOs: 10, 14 and 44.

11. A method of *in vivo* downregulating heparanase activity comprising the step of *in vivo* administering the antisense oligonucleotide of claim 8.

12. A pharmaceutical composition comprising the antisense oligonucleotide of claim 8 and a pharmaceutically acceptable carrier.

13. A ribozyme comprising the antisense oligonucleotide of claim 8 and a ribozyme sequence.

14. An antisense nucleic acid construct comprising a promoter sequence and a polynucleotide sequence directing the synthesis of an antisense RNA sequence of at least 10 bases being hybridizable *in vivo*, under physiological conditions, with a portion of a polynucleotide strand encoding a polypeptide having heparanase catalytic activity.

15. The antisense nucleic acid construct of claim 14, wherein said polynucleotide strand encoding said polypeptide having heparanase catalytic activity is as set forth in SEQ ID NOs: 9, 13, 42 or 43.

16. The antisense nucleic acid construct of claim 14, wherein said polypeptide having heparanase catalytic activity is as set forth in SEQ ID NOs: 10, 14 or 44.

17. A method of *in vivo* downregulating heparanase activity comprising the step of *in vivo* administering the antisense nucleic acid construct of claim 14.

18. A pharmaceutical composition comprising the antisense nucleic acid construct of claim 14 and a pharmaceutically acceptable carrier.

19. A nucleic acid construct comprising a polynucleotide sequence functioning as a promoter, said polynucleotide sequence is derived from SEQ ID NO:42 and includes at least nucleotides 2535-2635 thereof or from SEQ ID NO:43 and includes at least nucleotides 320-420.

20. A method of expressing a polynucleotide sequence comprising the step of ligating the polynucleotide sequence into the nucleic acid construct of claim 19, downstream of said polynucleotide sequence derived from SEQ ID NOs:42 or 43.

21. A recombinant protein comprising a polypeptide having heparanase catalytic activity.

22. The recombinant protein of claim 21, wherein said polypeptide includes at least a portion of SEQ ID NOs:10, 14 or 44.

23. The recombinant protein of claim 21, wherein the protein is encoded by a polynucleotide hybridizable with SEQ ID NOs: 9, 13, 42, 43 or a portion thereof at 68 °C in 6 x SSC, 1 % SDS, 5 x Denharts, 10 % dextran sulfate, 100 µg/ml salmon sperm DNA, and ³²p labeled probe and wash at 68 °C with 3 x SSC and 0.1 % SDS.

24. The recombinant protein of claim 21, wherein the protein is encoded by a polynucleotide at least 60 % identical with SEQ ID NOs: 9, 13, 42, 43 or portions thereof as determined using the Bestfit procedure of the DNA sequence analysis software package developed by the Genetic Computer Group (GCG) at the university of Wisconsin (gap creation penalty - 12, gap extension penalty - 4).

25. A pharmaceutical composition comprising, as an active ingredient, the recombinant protein of claim 21.

26. A method of identifying a chromosome region harboring a heparanase gene in a chromosome spread comprising the steps of:

- (a) hybridizing the chromosome spread with a tagged polynucleotide probe encoding heparanase;
- (b) washing the chromosome spread, thereby removing excess of non-hybridized probe; and
- (c) searching for signals associated with said hybridized tagged polynucleotide probe, wherein detected signals being indicative of a chromosome region harboring a heparanase gene.

27. A method of *in vivo* eliciting anti-heparanase antibodies comprising the steps of administering a nucleic acid construct including a polynucleotide segment corresponding to at least a portion of SEQ ID NOs:9, 13 or 43 and a promoter for directing the expression of said polynucleotide segment *in vivo*.

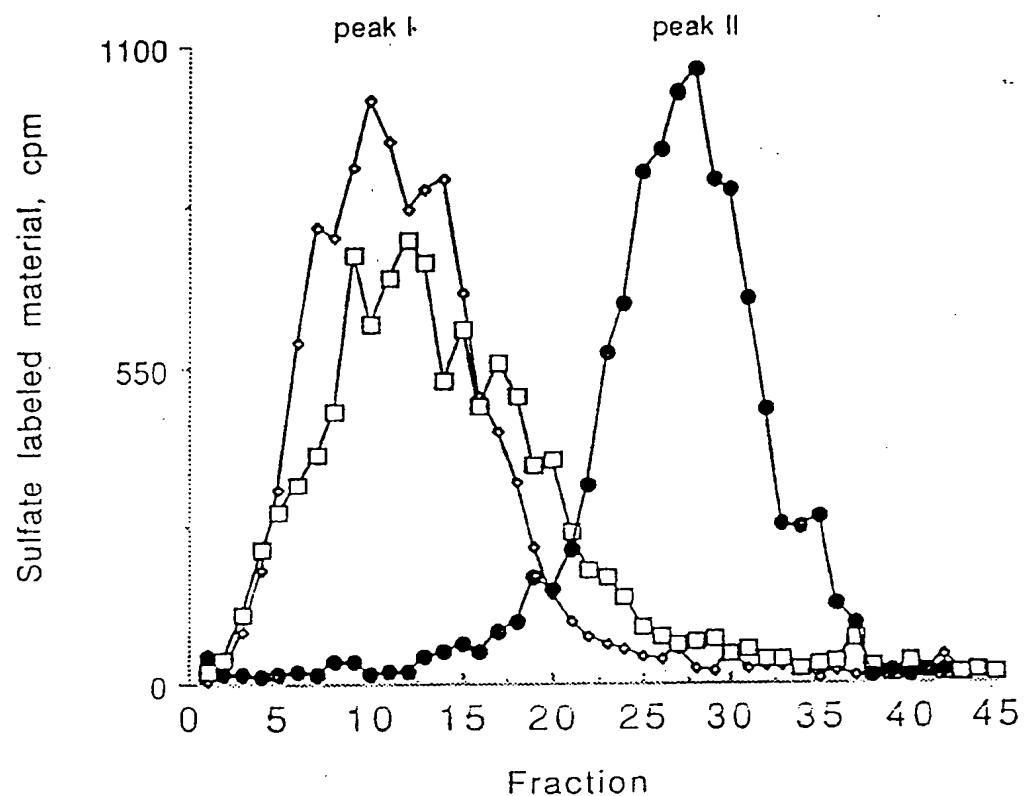
28. A DNA vaccine for *in vivo* eliciting anti-heparanase antibodies comprising a nucleic acid construct including a polynucleotide segment corresponding to at least a portion of SEQ ID NOs:9, 13 or 43 and a promoter for directing the expression of said polynucleotide segment *in vivo*.

1/34

Fig 1

1 CTAGAGCTTCGACTCTCCGCTGCGCGGCACTGGCGGGGGAGCAGCCAGGTGAGCCCA
 61 AGATGCTGGCTGGCTCGAAGGCTGGCTGCGCTGCCGCCCGCTGATGCTGCTGCCCTGGGC
 M L L R S K P A L P P P L M L L L G C P
 121 CGCTGGGTCCCCCTCTCCCCCTGGCGGCCCTGCCGCCCGCTGATGCTGCTGCCCTGGGC
 L G P L S P G A L P R P A Q A Q D V V D
 181 ACCTGGACTTCCTCACCCAGGAGCGCGCTGCACCTGGTGAGGCCCTCGTCCCTGCCGTCA
 L D F F T Q E P L H L V S P S F L S V T
 241 CCATTGACGCCAACCTGGCCACGGACCCCGCGGTTCTCATCCCTGGGTTCTCCAAAGC
 I D A N L A T D P R F L I L ' G S P K L
 301 TTCTGACCTGGCCAGAGCTTGCTCTGGCTACCTGAGGGTTGGTGCCACCAAGACAG
 R T L A R G L S P A Y L R F G G T K T D
 361 ACTTCCTAATTTCGATCCCAGAAGGAATCAACCTTGAAGAGAGAAGTTACTGGCAAT
 F L I F D P K K E S T F E R S Y W Q S
 421 CTCAGTCACCAAGGATATTGCAAAATGGATCCATCCCTCTGATGAGGAGAAGT
 Q V N Q D I C K Y G . S I P P D V B E K L
 481 TACGGTTGGAATGGCCCTACCAGGAGCAATTGCTACTCCGAGAACACTACCAAGGAAAGT
 R L E W P Y Q E Q L L L R E H Y Q K K F
 541 TCAAGAACAGCACCTACTCAAGAAGCTCTGTAGATGTGCTATACACTTTGCAACTGCT
 K N S T Y S R S S V D V L Y T F A N C S
 601 CAGGACTGGACTTGATCTTGCCCTAAATGGCTTATTAAGAACAGCAGATTGCACTGG
 G L D L I F G L N A L L R T A D L Q W N
 661 ACAGTTCTAATGCTCAGTTGCTCTGGACTACTGCTCTCCAAGGGGTATAACATTTCTT
 S S N A Q L L L D Y C S S K G Y N I S W
 721 GGGAACTAGGCAATGAACTAACAGTTCTTAAAGAAGGCTGATATTTCTCATCAATGGGT
 E L G N E P N S F L K K A D I F I N G S
 (T)
 781 CCGACTTGGAGAAGATTATATTCAATTGCAAAACTCTAAAGAAAGTCCACCTCAAAA
 Q L G E D Y I Q L H K L L R K S T F K N
 (F)
 841 ATGCAAAACTCTATGGCTCTGATGTTGGTCAGGCTCGAAGAAAGACGGCTAAGATGCTGA
 A K L Y G P D V G Q P R R K T A K M L K
 901 AGAGCTTCTGAAAGGCTGGTGGAGAAGTGTATTGATTCACTGCTTACATGGCATCACTACTATI
 S F L K A G G E V I D S V T W H H Y Y L
 961 TGAATGGACGGACTGCTACCAAGGGAGATTCTAAACCTGATGTATTGGACATTTC
 N G R T A T R E D F L N P D V L D I F I
 1021 TTTCATCTGTGCAAAAAGTTTCCAGGTGGTGGAGAGCACCGGCTGGCAAGAAGGTCT
 S S V Q K V F Q V V E S T R P G K K V W
 1081 GGTTAGGAGAACAAAGCTCTGCATATGGAGGGGGAGCTGGCTATCCGACACCTTGC
 L G E T S S A Y G G G A P L L S D T F A
 1141 CAGCTGGCTTATGGCTGGATAATTGGGCTGTCAGCCGAATGGAAATAGAACGG
 A G F M W L D K L G L S A R M G I S V V
 1201 TGATGAGGCAAGTATTCTTGAGGAGGAAACTACCATTTAGTGGATGAAACACTGGATC
 M R Q V F F G A G N Y H L V D E N F D P
 1261 CTTTACCTGATTATGGCTATCTCTGTTCAAGAAATTGGGGCACCAGGTGTTAA
 L P D Y W L S L L F K K L V G T K V L M
 1321 TGCAAGCGTGCAGGTTCAAAGAGAAGGAAGCTCGAGTATACCTTCATGCAACAAACA
 A S V Q G S K R R K L R V Y L H C T N T
 1381 CTGACAATCCAAGGTATAAAAGAGGAGATTTAACCTGCTATGCCATAACCTCCATAACG
 D N P R Y K E G D L T L Y A I N L H N V
 1441 TCACCAAGTACTTGGCTTACCTATCCCTTCTAACAGCAAGTGGATAAAACCTTC
 T K Y L R L P Y P F S N K Q V D K Y L L
 1501 TAAGACCTTGGGACCTCATGGATTACTTCCAAATCTGCTCAACTCAATGGCTCAACTC
 R P L G P H G L L S K S V Q L N G L T L
 1561 TAAAGATGGGGATGATCAAACCTTGCCACCTTAAATGGAAAAACCTCTGGCCAGGAA
 K M V D D Q T L P P L M E K P L R P G S
 1621 GTCACTGGCTTGGCTGAGCTTCTCATATAGTTTGTGATATAAGAAATGCCAAAGTGC
 S L G L P A F S Y S F F V I R N A K V A
 1681 CTGCTTGACATCTGAAAATAAAATACTAGTCTGACACTG
 A C I

FIG. 2



3/34

FIG. 3A

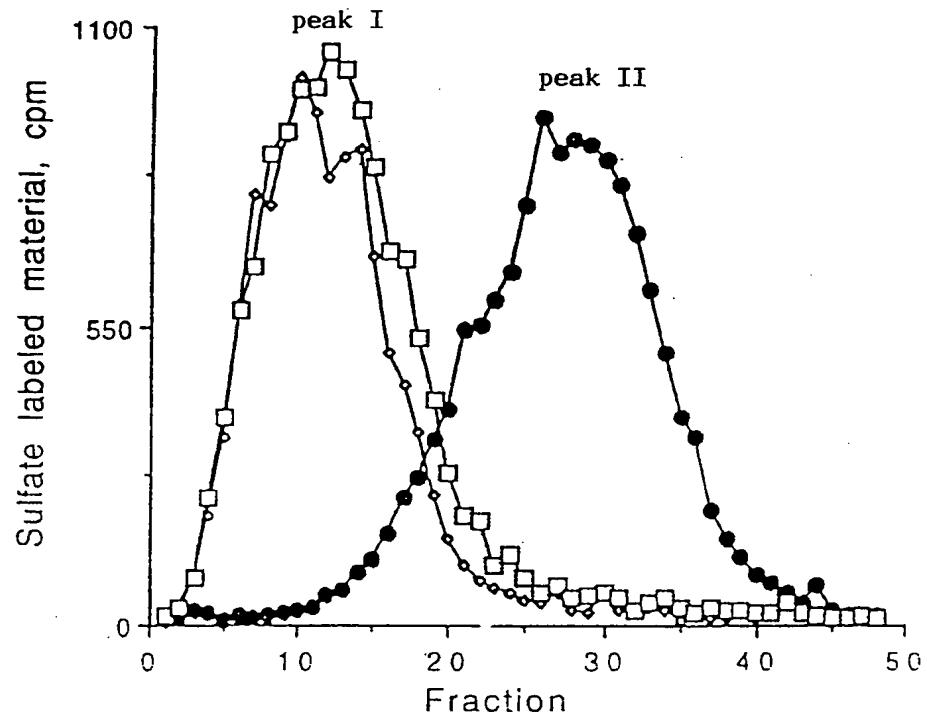


FIG. 3B

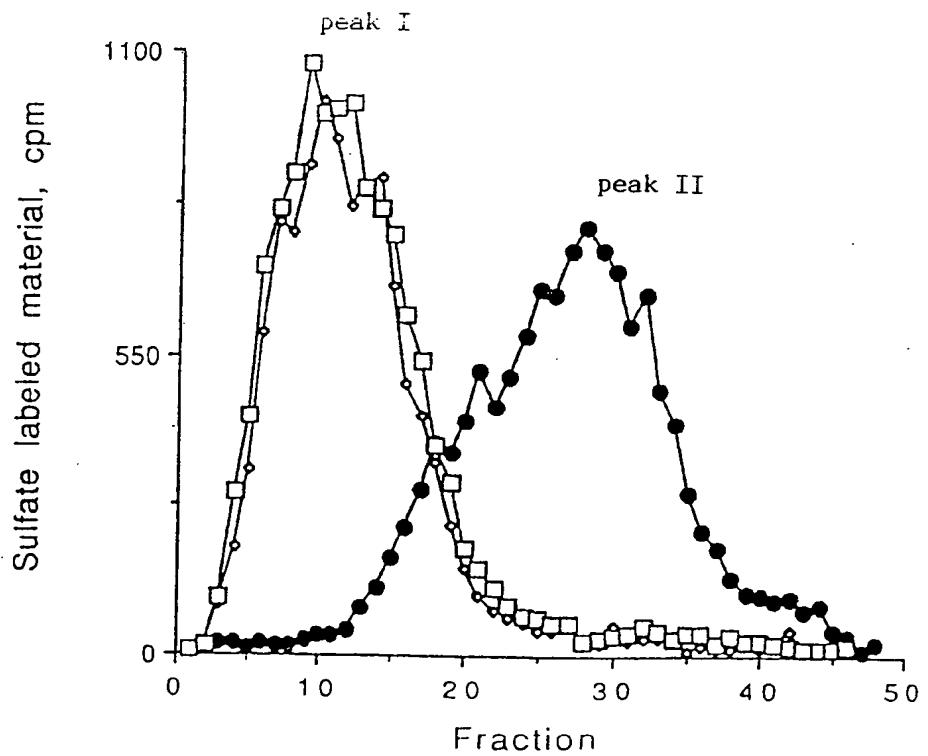
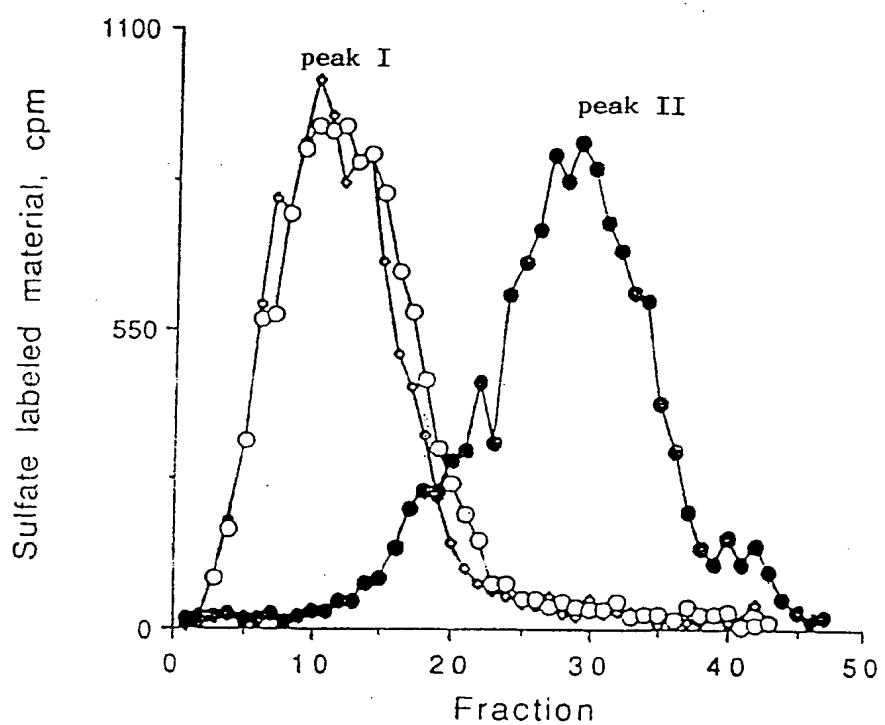


FIG. 4



5/34

FIG. 5A

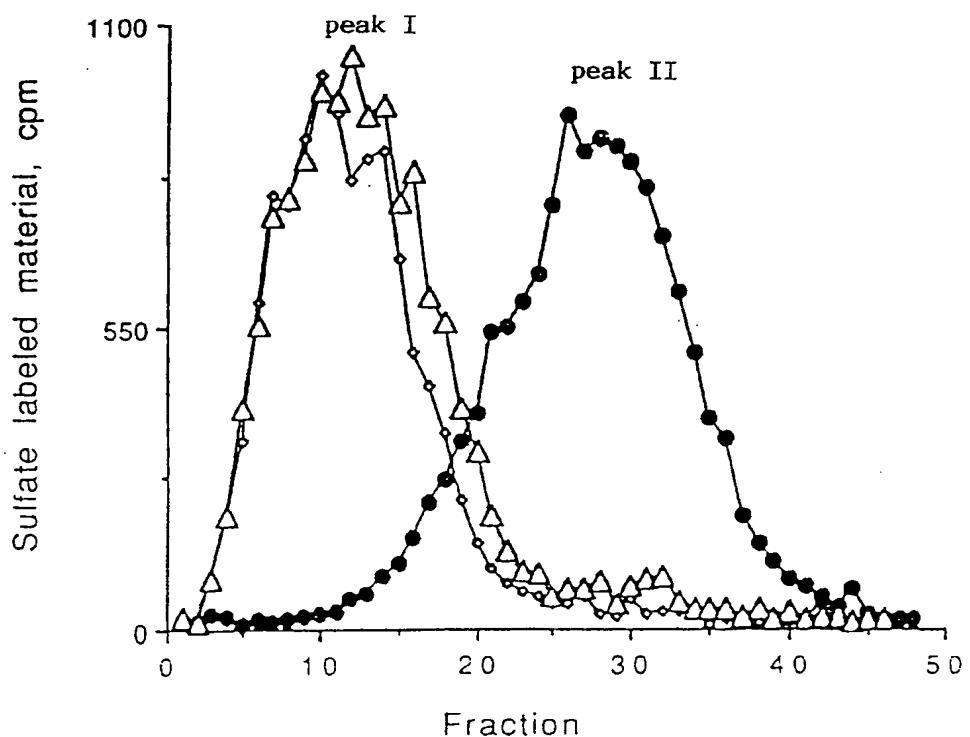
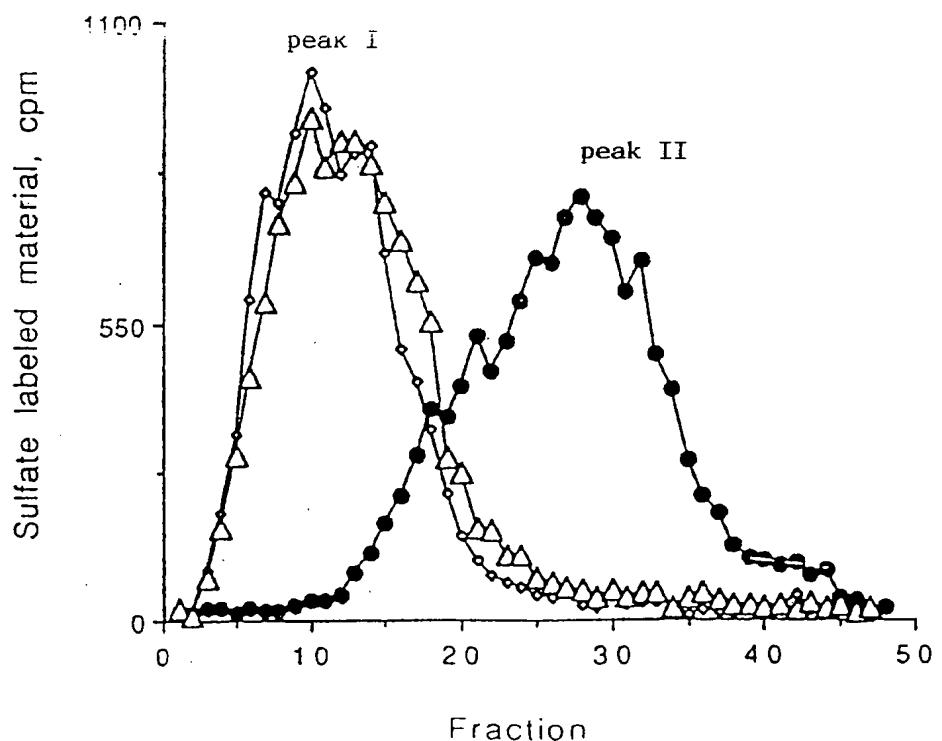


FIG. 5B



6/34

FIG. 6A

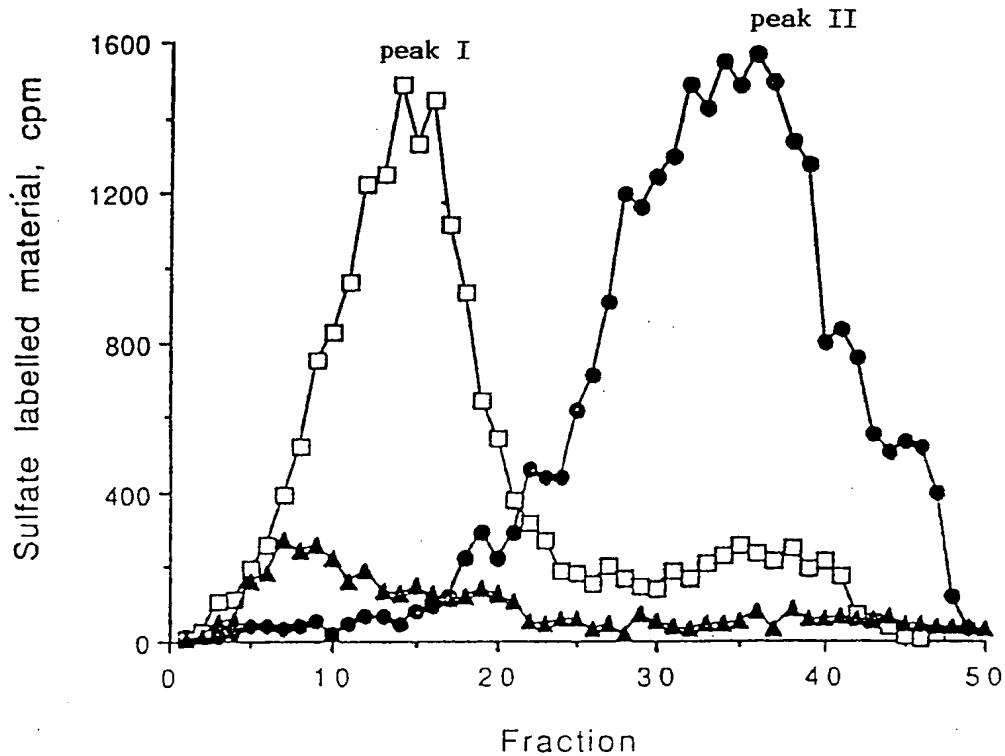
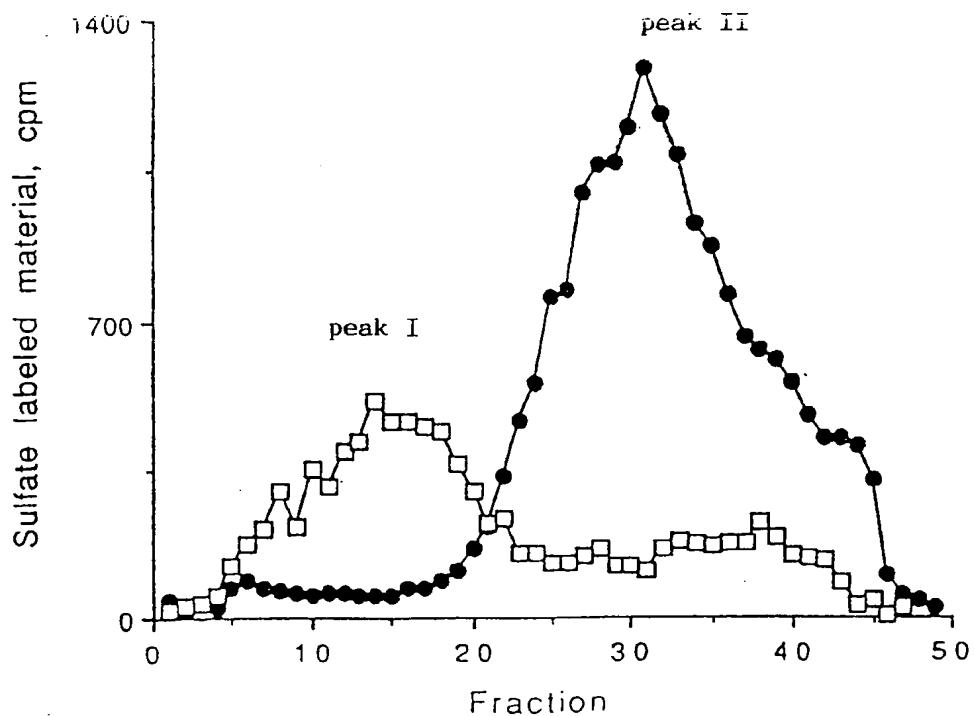


FIG. 6B



7/34

FIG. 7A

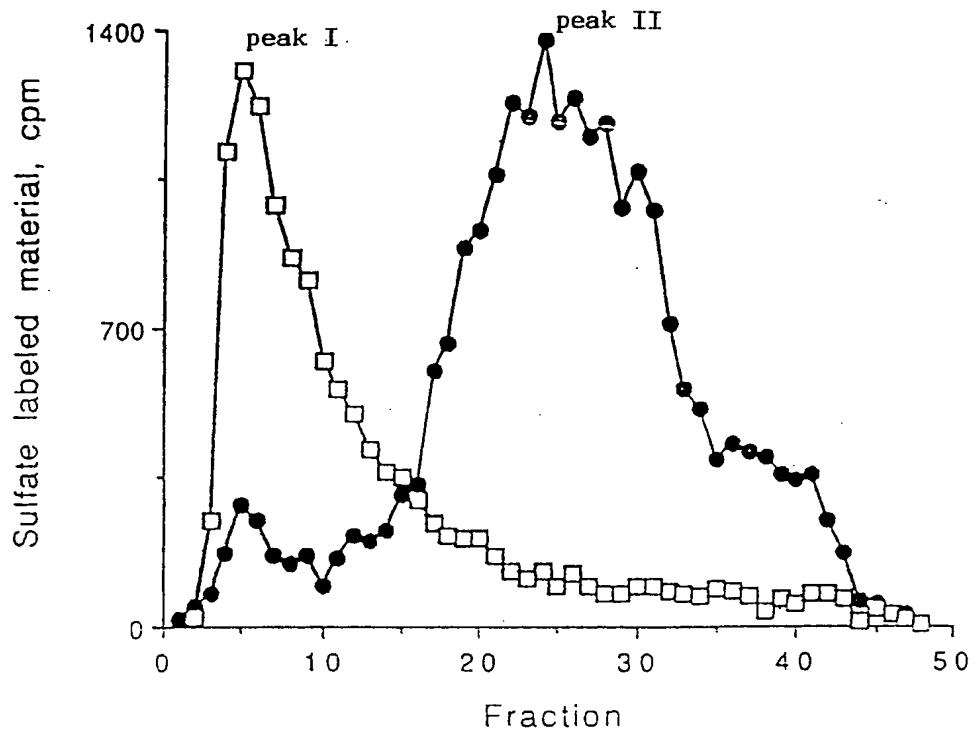
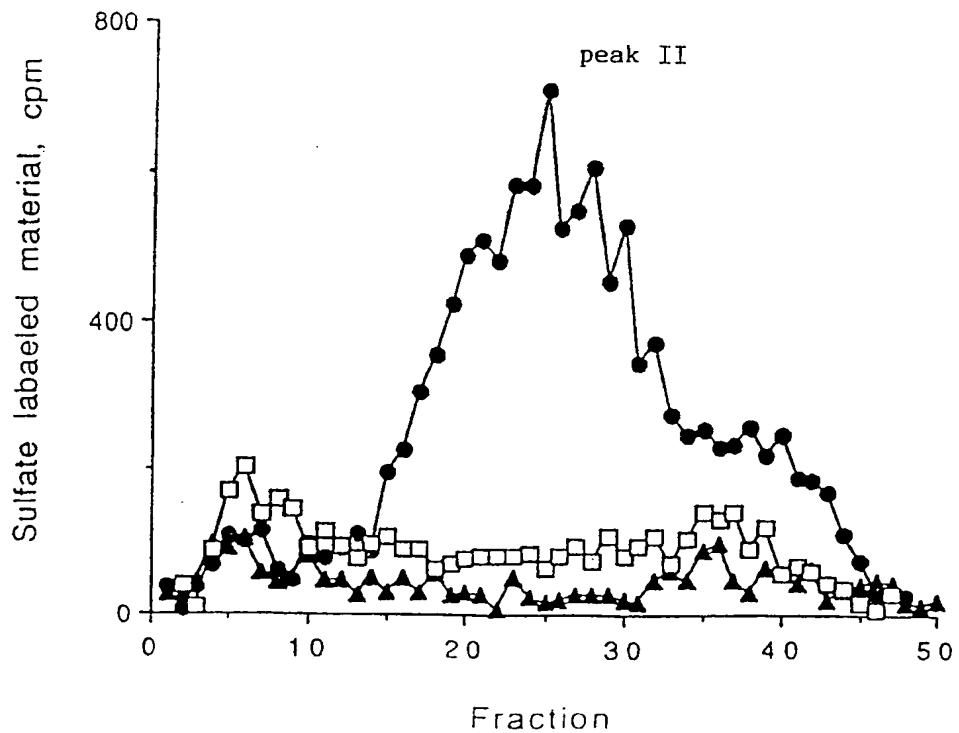


FIG. 7B



8/34

FIG. 8A

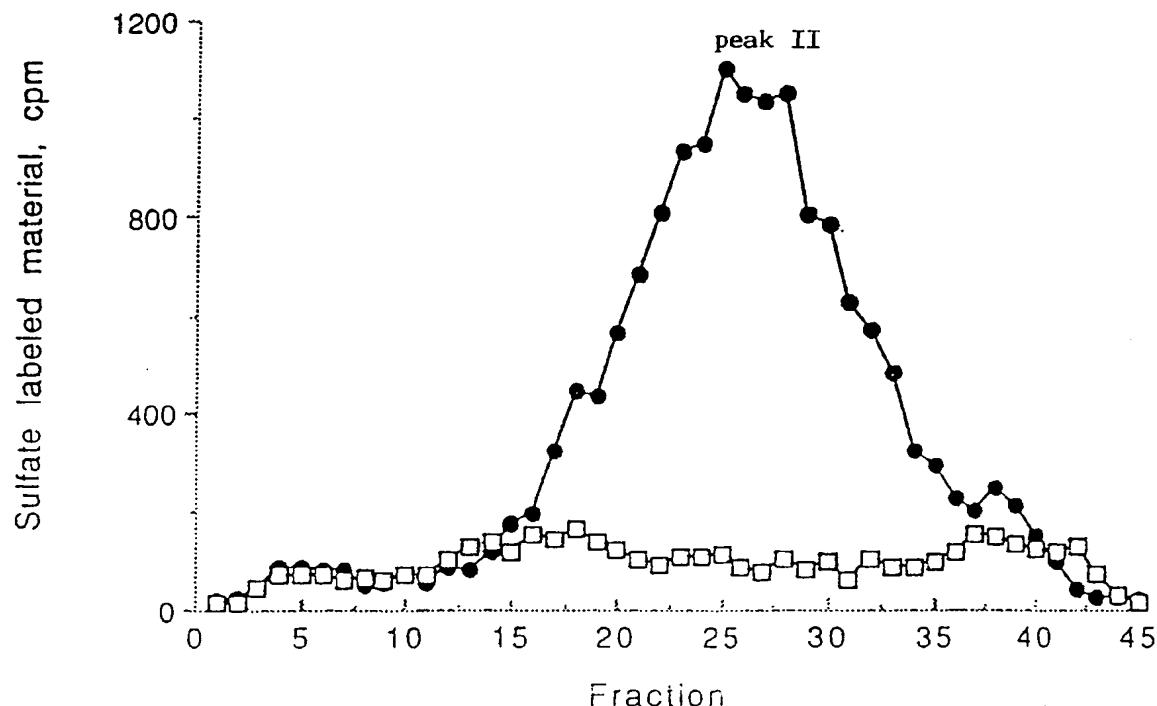
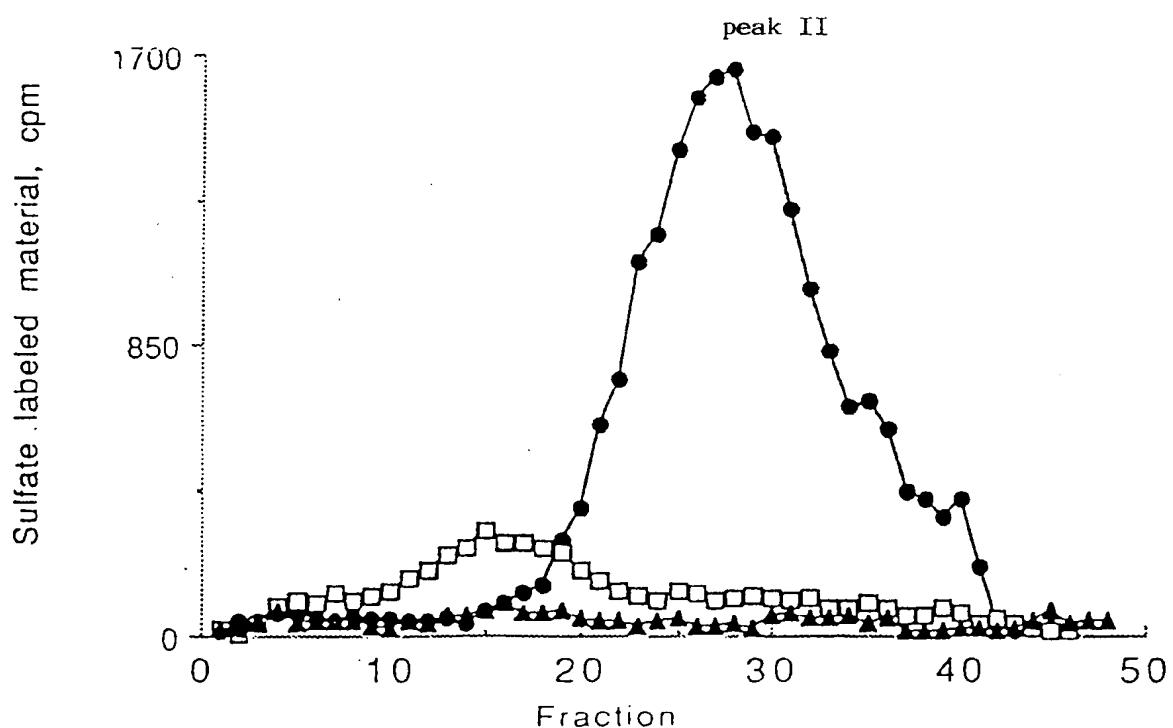


FIG. 8B



9/34

FIG. 9A

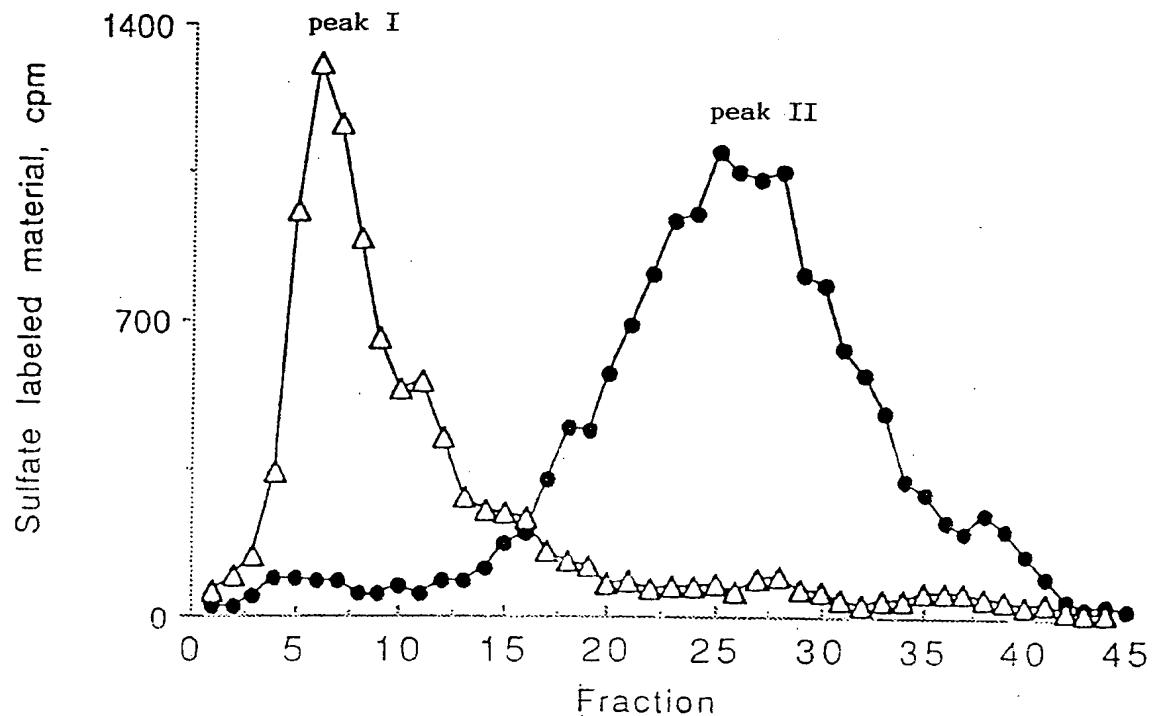
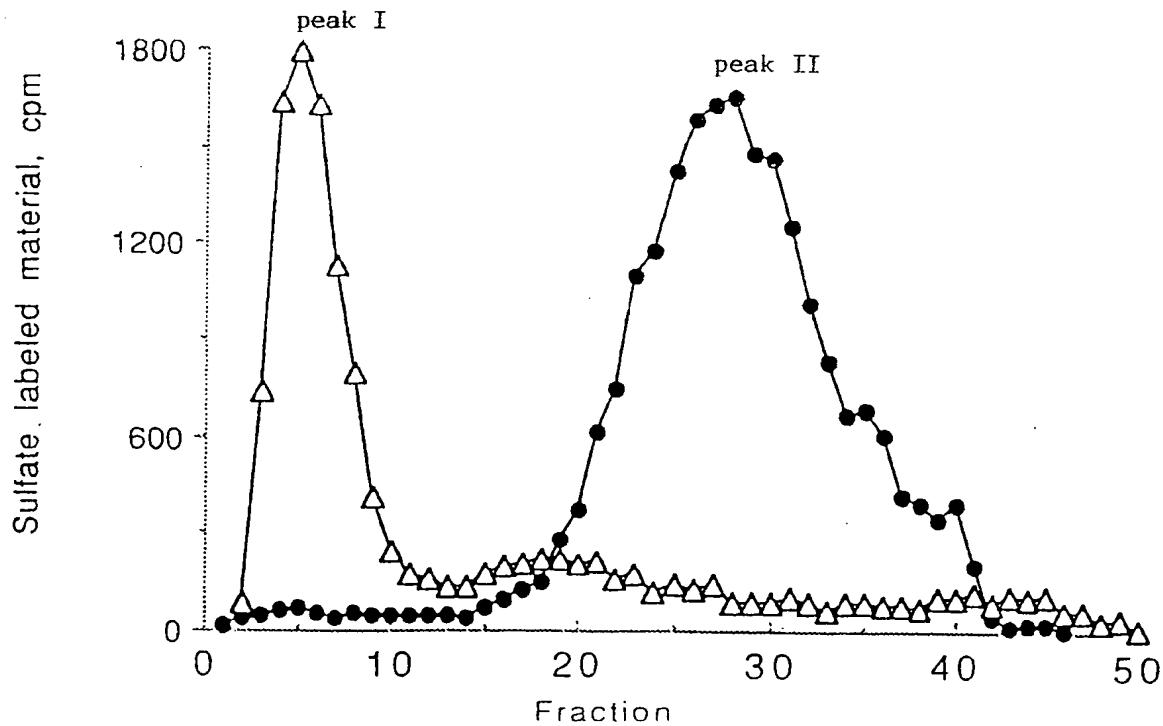


FIG. 9B



10/34

FIG. 10A

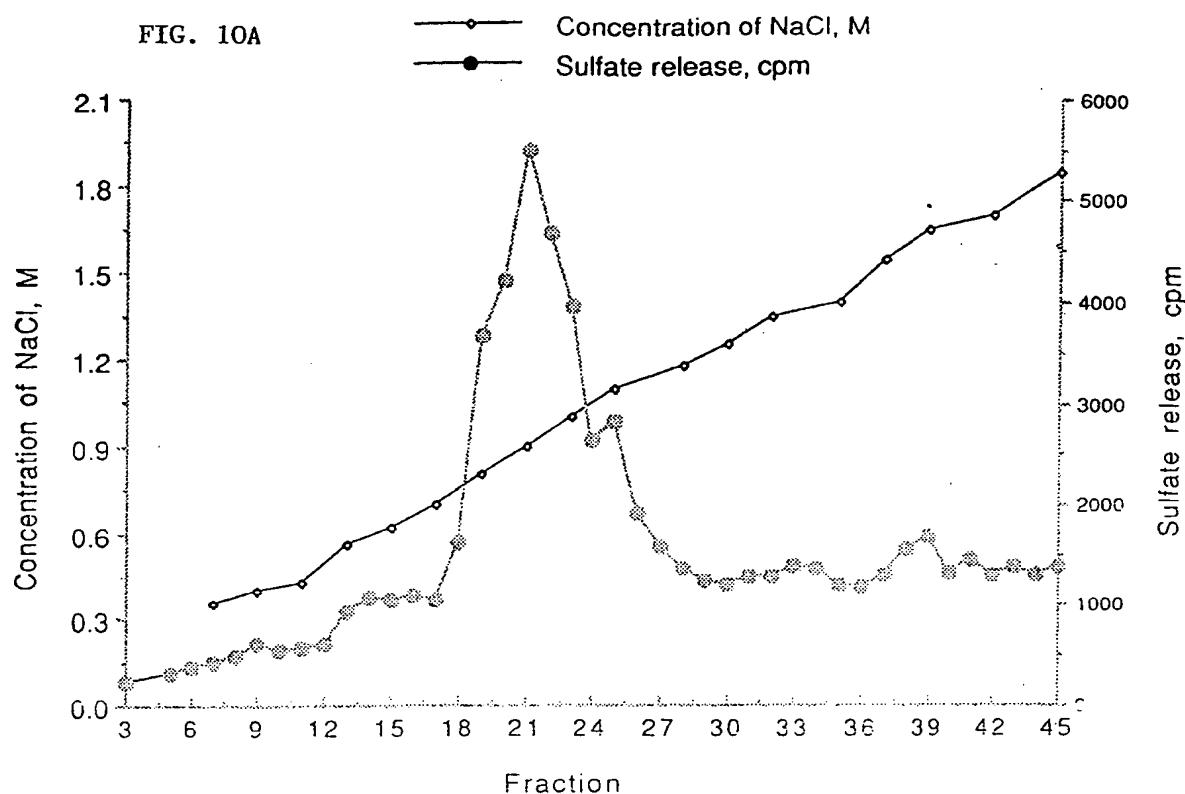
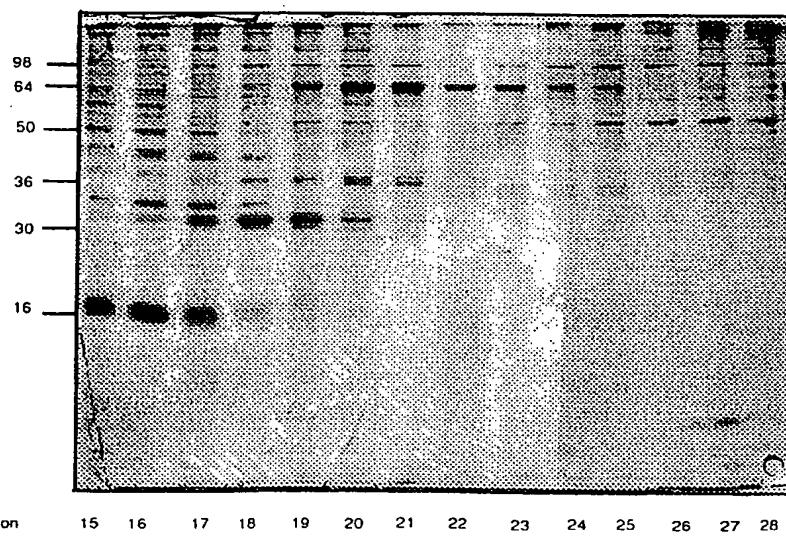


FIG. 10B

Heparin-avidogel

kDa



11/34

FIG. 11A

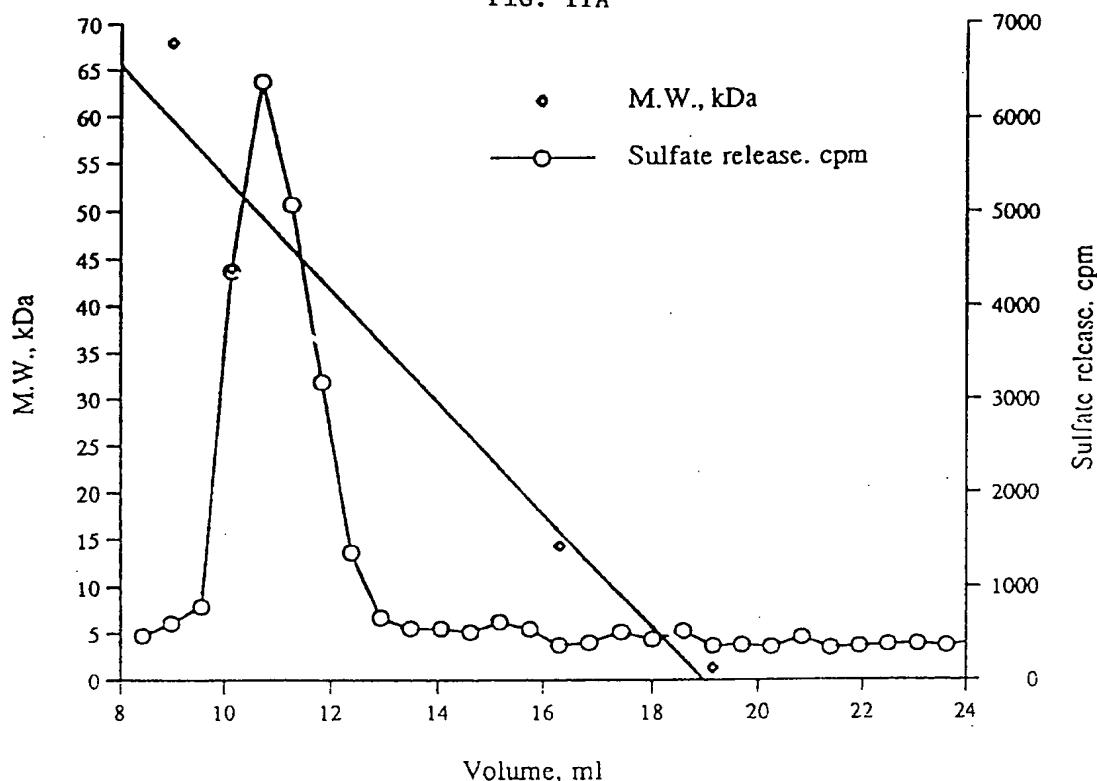
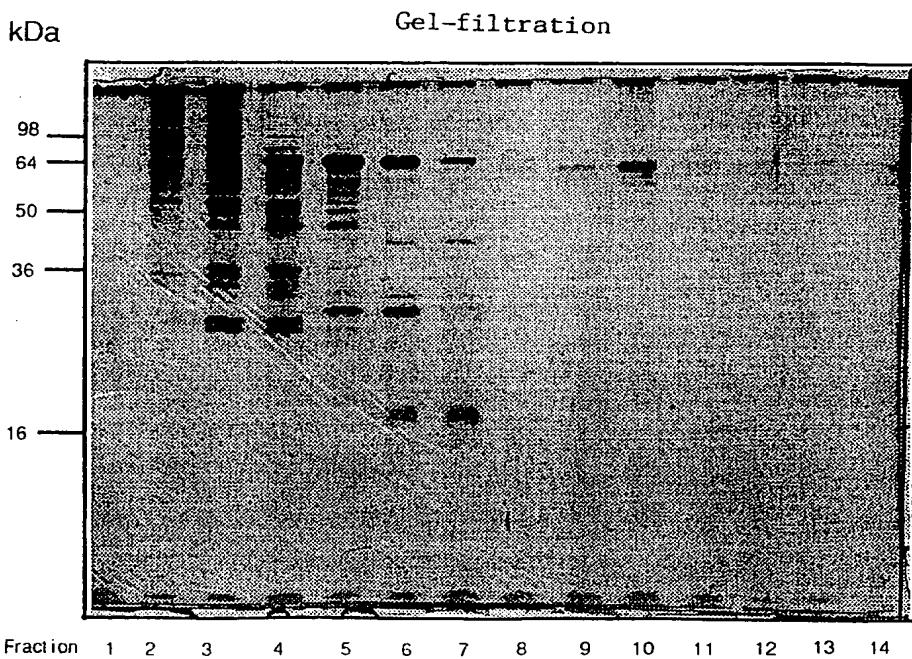


FIG. 11B



12/34

FIG. 12A

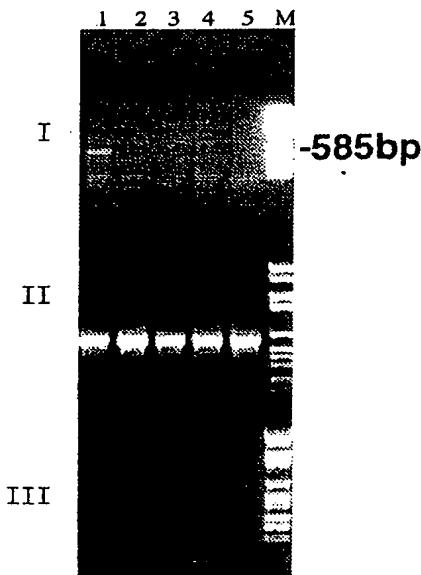


FIG. 12B

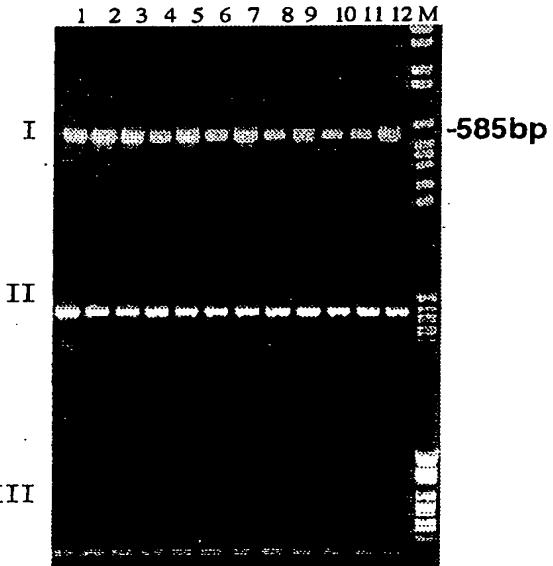


FIG. 12C

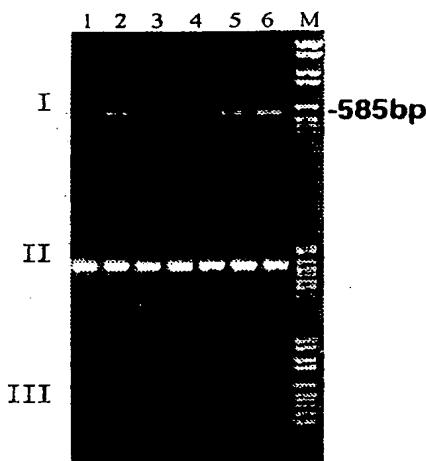


FIG. 12D

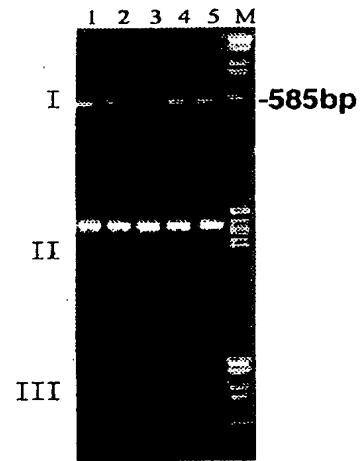


FIG. 12E

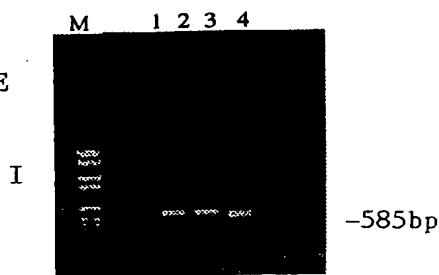


Fig. 13

mouse	CTGGCAAGAAGGTCTGGTTGGGAGAGACGAGCTCAGCTTACGGTGGCGGT	5Q
human	CTGGCAAGAAGGTCTGGTTAGGAGAAACAAAGCTCTGCATATGGAGGCAGA	1115
mouse	GCACCCCTGCTGTCCAACACCTTGCAGCTGGTTATGTGGCTGGATAA	100
human	GCGCCCTGCTATCCGACACCTTGCAGCTGGTTATGTGGCTGGATAA	1165
mouse	ATTGGGCCTGTCAGCCCAGATGGCATAGAAGTCGTGATGAGGCAGGTGT	150
human	ATTGGGCCTGTCAGCCCAGATGGGAATAGAAGTGGTGTGAGGCAAGTAT	1215
mouse	TCTTCGGAGCAGGCAACTACCACTTAGTGGATGAAAACTTGAGCCTTA	200
human	TCTTGAGCAGGAAACTACCATTAGTGGATGAAAACTTGATCCTTA	1265
mouse	CCTGATTACTGGCTCTCTCTGTTCAGAAAACGGTAGGTCCCAGGGT	250
human	CCTGATTATTGGCTATCTCTGTTCAGAAAATTGGTGGCACCAAGGT	1315
mouse	GTTACTGTCAAGAGTGAAGAGCCCAGACAGGAGCAGACTCCGAGTGTATC	300
human	GTAAATGGCAAGCGTCAAGGTTCAAAGAGAAGGAAGCTCGAGTATACC	1365
mouse	TCCACTGCACAAACGTCTATCACCCACGGATATCAGGAAGGAGATCTAACT	350
human	TTCAATTGCACAAACACTGACAATCCAAGGTATAAGAAGGGAGTTAACT	1415
mouse	CTGTATGCTCTGAACCTCCATAATGTCAACCAAGCACTTGAAGGTACCGCC	400
human	CTGTATGCCATAAACCTCCATAACGTCAACCAAGTACTTGCCTTACCTA	1465
mouse	TCCGTTGTCAGGAAACCAAGTGGATACGTACCTTCTGAAGCCTTCGGGC	450
human	TCCTTTCTAACAAAGCAAGTGGATAAACCTTCTAACGACCTTGGAC	1515
mouse	CGGATGGATTACTTCCAAATCTGTCCAACGTGAAACGGTCAAATTCTGAAG	500
human	CTCATGGATTACTTCCAAATCTGTCCAACCTAACGGTCTAACTCTAAAG	1565
mouse	ATGGTGGATGAGCAGACCCCTGCCAGCTTGACAGAAAAACGTCTCCCCGC	550
human	ATGGTGGATGATCAAACCTTGCACCTTAAATGGAAAAACCTCTCCGGCC	1615
mouse	AGGAAGTGCACTAAGCCTGCCCTTCTATGGTTTTGTCTAA	600
human	AGGAAGTCACTGGGCTTGCAGCTTCTCATATAGTTTTGTATAA	1665
mouse	GAAATGCCAAAATCGCTGCTTGTATATGAAAATAAA	637
human	GAAATGCCAAAAGTTGCTGCTTGCATCTGAAAATAAA	1702

FIG. 14

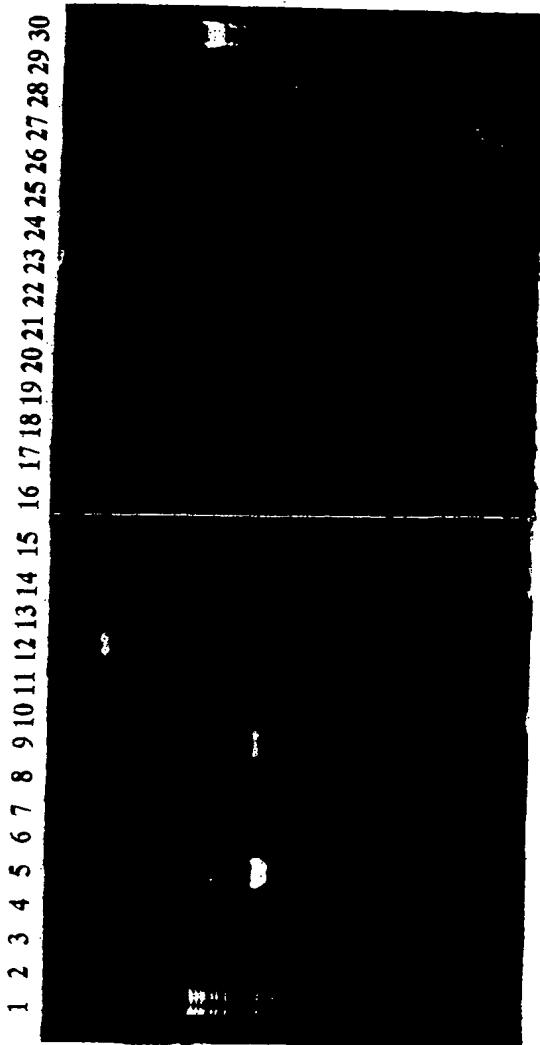


Figure 15

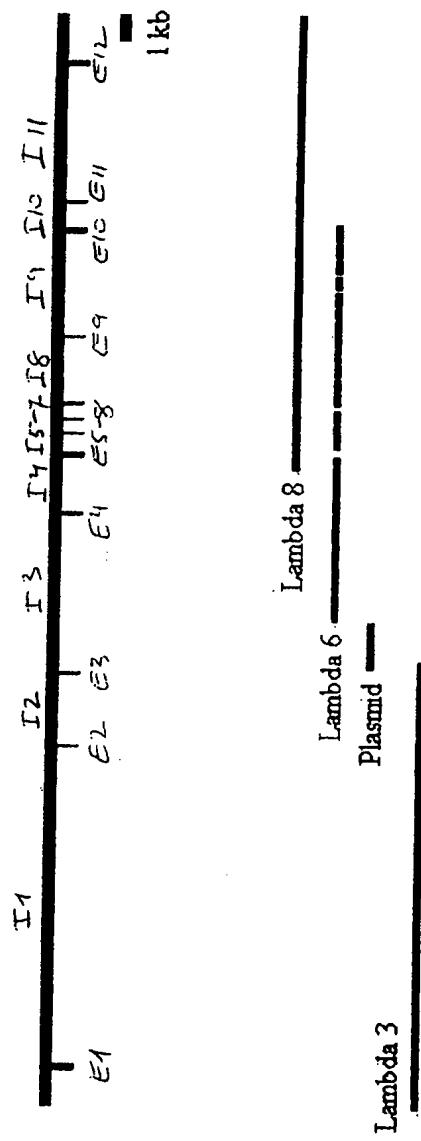


Figure 16

ggatctggctcaactgcaatctgcctccatgcaattcttatgcatca
 gcctccttagtagcttggattataggctgcgcaccactcctggctaca
 ccatgttgcggcaggctgtctgaactctggctctagtgatccacccg
 ccttggccctccaaagtgctggattacaggtgtgagccatcacccgg
 cccccggttccatattagtaactcacatgttagaccacaaggatgcacta
 ttagaaaaacttgcataatggccactttccaaatccacccaaacatgttaaa
 gaaattggtagactggcatggcagctggcatgcctgcataatcttag
 cattttgtgaggctgagacgggcagatcagggcaggatggaggatggagcc
 atctcgacagacatgtgtaaatcccatactctactaaaaataaaaaacta
 tagccgggggtgatggcaggcccgttagtcccagctactcgggaggctg
 aggcaggagaatggcgtgaatccaggaggcagagcttgcagtgagccgag
 atggtgcactgcactccagctggcgcacagagcgcagactccgtctcaa
 aaaaaaaaaaaaaaaaagaaaaattggtagactgttgcataacacaggag
 tcagggcatgggtgggtgtaaagattaatgtcatgacaaatgtggaaa
 agaaaacttctgttttccaaactccacgtctgttacccatattacactc
 ttctggtagtgtgtgtttatgtgtaaatttttcatatgtatacagt
 aattgttaggatatacgttgcattctactatgacggatgttgcata
 gcttttaagtgcataagaataggtagatctatgcaaaactcaactatgac
 ttattttatatttaagagagggtctacttgcacccaggctggatgc
 agtggtagattaggctactgcaccccccaggctcaaaataa
 acctccacccatcagcccccaggtagctgaaaccacaggcgcacggccacc
 acgcctgtcaattttttgtatgttagatgggggttcatatgt
 tgcccaggctgttgcattcccggtcaagcaatcccccacccctgg
 cctcccaaaatgtggcatcacaggcatgtggcatcaactggcatcacat
 accatgcctggctgatttatgcaatttagatatgcatttcaaaaataatc
 tattttatgttgcatttgcatttgcataatctcaagtggaaaatct
 aagggtttgggttattgttgcatttgcatttgcataatatttagacttca
 ctaagcaccaacatgatcacatgcctgagctatggctagcatgcgtgt
 agacaaaacttaatctgtttggtagatataatcttagatgt
 ccaatgttagcaacatcacaataactaacaatttagagatgtca
 tggctcaacaaattggaggatgtacagagatgtcaacaaattggaggatgc
 tatgagatgttgcattggagagctgcctggagatgttagagaaaaatgt
 tgaggaaagtacatttcagctaaacacactgcctatgtcgagg
 tgtaactgcattcacatcccgattctgcacacttcacatcccgattctgac
 acttcacccaggtaactgtctcagagcttgggtccgcattgtgt
 gagactatgcacttggcagggtgtgagaagggaaagagaacacaagtaaa
 gcacctgtatcaggcatacagtaggcactaagcgtgcattgcatttgc
 attatacatcagtgtaagcatcaaggaaaagctgaagaaaagtctgacca
 acagcggaaagataaatgcgcagaggagaaatttgcataaggctccaaatt
 cagggcaggctgtactctacactttgtatggggctcaggctgt
 tccagacatttgcaggcaactaaccccttaagattgtctaaatattgtcttaa
 tgagaaggtagataaaaatgggtgttgcatttccaggctgc
 tttagcgtatgtctgaggccaggattttcaagcaaaaaggtaaaaatctgag
 aaactgcctggccaggaggacaatcagattttggctcaagtgc
 caagtgttataaaggatgtatggggatggatgaaatactccatgg
 gggtttactcgcgggttgcaggatcccgccatcagaatgggatct
 gggagtccggaaacgcgtgggtccacagagcgcgcagaacacgc
 aggaaggctgttgcgggatgcaggctgtctcccgccgc
 gggcgtctcccccaggctccggcgcttggatcccgccatctccgc
 acccttcaagtgggtgtgggtgatttcgttaagtgaacgtgacc
 agggggaaagcgcaggcaagggatgttagggagagagccgg
 tggtggatggggatgcaggatggggatgcaggaaagaggatgg
 gggcgcaggatggggatggggatgcaggatggggatgg
 AAGGGCGCTGGGGCTGGGGAGGAAGTGCTAGAGCTCTCGACTCTCCG
 CTGCGCGCAGCTGGGGGGAGCAGCCAGGTGAGCCCAAGATGCTGCT
 M L L
 GCGCTCGAAGCCTGCGCTGCCGCCGCTGATGCTGCTGCTCCGGGGC
 R S K P A L P P L M L L L G
 CGCTGGGTCCCCCTCTCCCCCTGGCGCCCTGCCCGACCTGCGCAAGCACAG
 2800
 2850

Fig. 16 (continued)

17/34

P L G P L S P G A L P R P A Q A Q	2900
GACGTCGTGGACCTGGACTTCTTCACCCAGGAGCCGCTGCACCTGGTGAG	
D V V D L D F F T Q E P L H L V S	
CCCCTCGTCTCCTGTCCGTACCATTGACGCCAACCTGGCACGGACCCGC	2950
P S F L S V T I D A N L A T D P	
GGTTCCATCCTCCTGGGtaagcgccagcctccctgtccctgtcccttt	3000
R F L I L L G	
tcctgtccctctgcacacccatgtctgccccggccagggctccctttttt	3050
tgcgcggaaaacaaacttcacacccggAACCTCCCGCTGTCCTCCCCACC	3100
ccactttccgccttcatttcctccctccctccctactctcagaaaa	3150
aaccgcgtttgggggtatcattaaaaaaatagatttaggggttacaag	3200
tgcagttctgttccatgggtatattgcattgtggcatctggcttctt	3250
agtgttaactgtcacccgaatgtgtacattgttatctaataaggtat	3300
catccctcatccctcccccacccctccacccctttggagtctccagtgct	3350
actattccactaagtccatgtgtacacattgttagcggccactctaaat	3400
gagccttttgggttatttcattctgtatgttgcattgttgcataaggcaccaccaa	3450
ggtcagggtataagtggaaatttggaaaaactgcccacttgccccagt	3500
acttccctagccaagaggaggaaaccaggcagggtcacctgaaggcctg	3550
tgagtgttgcattgtgtgcagtgttaggacaagtaagattgtgcata	3600
cttctgtatTTTAAAGACTGTGTTAGGAAGATTCTCTTCTTTCTTCT	3650
TTTCTTTCTTTCTTTCTTTCTTTCTTTCTTTCTTTCTTTCTTCTTCT	3700
cagaacaggaaaaaaatctaataattcaataatggacacttaggagact	3750
actgcagtgcattacaagtccataaaaagatgtctccaaaaatggg	3800
gctgcggaaatgtggctgccttatcagctctaatgtttttcccttac	3850
agaaaagaaggaaacctgtatgcagggttcagggtctgcggccatgatgc	3900
gctgactccaagatgggagctacagggacaatcccaaggcttcttagg	3950
tcttatttaggcctggagccctccagagatggccacatcttgcac	4000
cagatagagggaaagatcaccattatctcacctctgtgtcaataac	4050
atgctgtccctccctgagcccacactatagttgcagcgtcaattaaatgg	4100
gtagtgtaactggtaagagatggacagaccatctggcttgcactc	4150
tctggcaagatgtactgtttccatctcttgcacccaa	4200
ccttgatttcttcagctgttagaaatggaaatttctcaaggtgcctcaagg	4250
ttatgg	4300
catagtaagtgttgcgttcaacacgaaattgtttcttgcattggatcat	4350
tgagcatttgttagccattcacgggtttctgtttttgtcatagtt	4400
aacctctctttcttctggcaactacaatttctggggggaaatcc	4450
ttacttt	4500
ctagttgggggataggaagattttccagagaaaatgtgtcaaccatagg	4550
tccagatcacaggccccacttgcatttttttttttttttttttttttttt	4600
ggcggttactgaacatgggtatgaagtagatgtccattactgaaatgt	4650
gaggacctgaggcccttcttcttcttcttcttcttcttcttcttcttctt	4700
tcccaagaaaggacagatggggttccccctggatgtacagg	4750
aaaaaaacatacagtggacttccaggatctggccatcccccaaccc	4800
gtcaagtcggccattgtactaacaccccccttaacacgtaaattcca	4850
atctgcatt	4900
tcattttccagccatcccccttttttttttttttttttttttttttttttt	4950
gactggagttataatggaaatccaagagagacataacatgtgtata	5000
ctttttttctgggttacaaaacagttcatttttttttttttttttttttt	5050
tccaggctggctgtctgtttccagggcccttcgttgagggccat	5100
ctgccatcacctgtccccagacgcattgcacaaagcaca	5150
ctgctaaagacctaaaagaggaggaaaccccttcctcatctaa	5200
gcttcattaaatttaggtgtgggtccatctccctggggccacagg	5250
ccaaacagccagccatctcagaagacaactaagtttgcatttttttttt	5300
acagtagaggagactaagacgcctgttttaatttattacgttcc	5350
aaagtgaagatgtggggggggatggcaagagactgagcagacgaa	5400
aaaaataaaggaaagaggaggaggacacaaacagctgcacacttcc	5450
cttgcattttgcctggcccttttttttttttttttttttttttttttt	5500
tttagtcttgctacaacactgtgttagtaacttagtttttttttttttt	5550
aaaaatgaagaaagtgggggggggggggggggggggggggggggggg	5600
tgaaaactagactctgtatcacatgtatgtgcacccataaaaagg	5650
gcagattatatttttaaggaaagagagactgaggatatggtagaaaa	5700

Fig. 16 (continued)

18/34

19/34

Fig. 16 (continued)

Fig. 16 (continued)

20/34

Fig. 16(continued)

21/34

aaataaaaattaaacaatactgcctgtaatctcagctaccaggaggcag	14900
aagcagaactgcttcaaccggcaagcagaagttgcagtgaagttagatc	14950
gcccactgcactccacgcggaaacagagcaagattctgtctaaaaaa	15000
caaaaacatgtattccctctaagtcctgcacaggaaatgttaaga	15050
aataggccaccaggaaagaagaagtaagaatgtttactatgtct	15100
tggaaaaatagttatactttcttgcttgcctaaacagTTCTCCAAA	15150
S P K	
GCTTCGTACCTGGCCAGAGGGCTTGTCTCCTGCGTACCTGAGGTTGGTG	15200
L R T L A R G L S P A Y L R F G	
GCACCAAGACAGACTCCTAATTCGATCCCAAGAAGGAATCAACCTTT.	15250
G T K T D F L I F D P K K E S T F	
GAAGAGAGAAGTTACTGGCAATCTCAAGTCAACCAGGgtgaaaattttta	15300
E E R S Y W Q S Q V N Q	
aagattcacttatatttaattaacgtcagtccgtcatgagaatgcttt	15350
gagaaaaactgttatttctcacacctaacaattatgagatataacttcctc	15400
tccctctatctgacccgtggagaaatctgaaacaagaggaggaggcagtgg	15450
gcagggttccttcatatcatgtttgtcatgtttcgttgcgttgcgttgc	15500
aaaaaaaaaaaaaaaaaaaaaggccctgtggatataactcgatggacttgcatttgc	15550
tacagtaaatattaataaaaaacagtgattgtatgcgttgcgttgcatttgc	15600
tggagggagcaagtgggttagatccgttgcgttgcgttgcatttgc	15650
caaagacacaatgatagattgtggatatttattctaaatataatgt	15700
ggtgaacgagatctgtggacttctggctccaaacgttgcgttgcatttgc	15750
agcaagctgtcaggggattctgtatattgtggatatttgcatttgcatttgc	15800
agaaacctgccttagggggccatggaaatttgcatttgcatttgcatttgc	15850
ctatcagacatcaaafggaaatctgtatcttacaaattactaggat	15900
gggcgcagtgtactcacccctgtatcccaacactttggggctggca	15950
ggaggatacttgaggcccaggatccgtggaccagcgttgcatttgcatttgc	16000
agacgttgcatttgcatttttataattttaaagaaaaactactgtaaaaaa	16050
tattgtatcacccactgttgcattttataatgtgtatataatgttatatttc	16100
attatgagaaatattgttattttcatatattatattttgcatttgc	16150
tattttatccaggatgttgcattttgcattttgcatttgcatttgcatttgc	16200
taaatttgacagaatgttgcattttgcattttgcattttgcatttgcatttgc	16250
tctttgaataaccagggtggatgtttatgggttgcatttgcatttgc	16300
ggcggttttttagacagagtcgttgcatttgcatttgcatttgcatttgc	16350
ggcacaacatggccactgtcatccgttgcatttgcatttgcatttgcatttgc	16400
tcccaccttagccctgttgcatttgcatttgcatttgcatttgcatttgc	16450
ccctgttgcatttgcatttgcatttgcatttgcatttgcatttgcatttgc	16500
ggctgttgcatttgcatttgcatttgcatttgcatttgcatttgcatttgc	16550
aaagtattgggattataggcatgttgcatttgcatttgcatttgcatttgc	16600
tagacttggcccttccaccaggcatttgcatttgcatttgcatttgcatttgc	16650
tgttagacaggaaactgttgcatttgcatttgcatttgcatttgcatttgc	16700
aggggggatgttgcatttgcatttgcatttgcatttgcatttgcatttgc	16750
atccctctgtgagcccagggtggatgttgcatttgcatttgcatttgcatttgc	16800
tgtgcagcttccagaaaggaaactgttgcatttgcatttgcatttgcatttgc	16850
gccagcaccggcaaccctgttgcatttgcatttgcatttgcatttgcatttgc	16900
acaggatgttgcatttgcatttgcatttgcatttgcatttgcatttgcatttgc	16950
gtctggctgttgcatttgcatttgcatttgcatttgcatttgcatttgcatttgc	17000
tgcaggcttgcatttgcatttgcatttgcatttgcatttgcatttgcatttgc	17050
atgttgcatttgcatttgcatttgcatttgcatttgcatttgcatttgcatttgc	17100
tttttatgttgcatttgcatttgcatttgcatttgcatttgcatttgcatttgc	17150
cctgttgcatttgcatttgcatttgcatttgcatttgcatttgcatttgcatttgc	17200
caagcgttgcatttgcatttgcatttgcatttgcatttgcatttgcatttgcatttgc	17250
gttgcatttgcatttgcatttgcatttgcatttgcatttgcatttgcatttgcatttgc	17300
caatgttgcatttgcatttgcatttgcatttgcatttgcatttgcatttgcatttgc	17350
cactgttgcatttgcatttgcatttgcatttgcatttgcatttgcatttgcatttgc	17400
acccttaagagaactgttgcatttgcatttgcatttgcatttgcatttgcatttgc	17450
ttcccgatgttgcatttgcatttgcatttgcatttgcatttgcatttgcatttgc	17500
agtgttgcatttgcatttgcatttgcatttgcatttgcatttgcatttgcatttgc	17550
tccgcctgtgggttgcatttgcatttgcatttgcatttgcatttgcatttgcatttgc	17600
gtatgttgcatttgcatttgcatttgcatttgcatttgcatttgcatttgcatttgc	17650
tcctgacccatctgttgcatttgcatttgcatttgcatttgcatttgcatttgcatttgc	17700

Fig. 16 (continued)

22/34

Fig. 16 (continued)

23/34

Fig. 16 (continued)

24/34

tcaagtttaacttaggtgcctgcacttttagttgtctaattcctgttagctg	23650
tacccatgcattcactgggtgcctccagctgcctgcacagatggaa	23700
aaccatagtcctataactctaggccaatttttaatgtaaaatttggattc	23750
attttaaattaataaataaataacaggaatttttaatgtttaaaatttggaa	23800
tataattaaaattatcaaaaatatttttaactgaacttgtgactagagat	23850
attttagattatgaagagtggggtttatgctaactaatgacagtctggcta	23900
tgcatgtggagcactgagctataaattgtgcttcccaattctcctgat	23950
gtcacttgaacaaaacctaagtgtcagaccagagcttggtatcttcca	24000
tgggatttcattcaacagctggagcaatgaagtcagattgattttttt	24050
aatttgcacatttttgtgtctcaaaaacataattataatcatttattag	24100
aactagaatttcttcgtttaacacacagaatagtattcattatgaaaa	24150
gcgaatctggaggcccttattgtgtgcctaaaccattaaatttgcata	24200
cgttttcttttagGAAGCTCTGTAGATGTGCTATACACTTTGCAAAC	24250
R S S V D V L Y T F A N	
GCTCAGGACTGGACTTGATCTTGGCTAAATGCGTTATTAAGAACAGCA	24300
C S G L D L I F G L N A L L R T A	
GATTTGCAGTGGAACAGTCTAAATGCTCAGTTGCTCCTGGACTACTGCTC	24350
D L Q W N S S N A Q L L L D Y C S	
TTCCAAGGGGTATAACATTTCTGGAACTAGGCAATGgtgatccccca	24400
S K G Y N I S W E L G N	
gggaacaattcattaataaggagattccccactagcattatttctttct	24450
tttcttt	24500
tgcccaggctggagtgcagtggccacccctggctcaacttgaagctctgc	24550
ctcccaaaacgcacattctcctgcctcagcctcccgagtagctgggactac	24600
aggcacccgcaccgcgccccgtatatttttttttttttttttttttttt	24650
ttttttgtattttagtagagacgggtttcaccgttttagccaggatg	24700
gtcttgatctcctgacccctgtatctgcctcctcgccctccaaagtgc	24750
tgggattacaggcgtgagccaccaggccccgttagcattatttcttatga	24800
cacttt	24850
agtgcagtggccatctcgctactgcacgtccacccctccagggttca	24900
cgccatttcctgcctcgcctcccgagtagctggactacacgcaccccg	24950
ccaccacgcggcgtatattttttgtattttttagtagagacgggtttca	25000
ccgtgttagccaggatgtctctatccgcacccatgtatctgcccggcc	25050
tcggcctccaaagtgtgggattacaggcgtgagccactgcgccccggcc	25100
aacactcttt	25150
caagtgcctcaacaatgcacacttttggaaagtgcacatggcagaaactctg	25200
ttgtatattttccagaaacctattattgtcaatcccagttatgttacatt	25250
tgaagtgagaaccaggatggagccacgcacgttcccagtcacgttccc	25300
ttgagatttcagaatcacttaaccctattatgttggcaaccctggactc	25350
agcaaaaactggaaagtgcacgttttttttttttttttttttttttttttt	25400
gtttctcaatgtgtcagttaatctcagtaacccatttgcacccatttca	25450
acctgcacccaaaggcgtctagaacttgcctgatagaatccatcgtgggtca	25500
agctcctgactgtctcttcttactctttttgtcaaaaactgtaaa	25550
tttttaactataatgttttttttttttttttttttttttttttttttttttt	25600
gtttttccacatcatcgcacatggaaataaggatataatgggaaatgaa	25650
atgttagtaataggataacgcacaaatgttttttttttttttttttttttt	25700
tt	25750
ggcgtgttcatacgctcaatgtaccccttgcacatgttttttttttttttt	25800
ctcacacccctcagccccctgattagctaggactacactatgcctagccaat	25850
tt	25900
ctcaagtaatccctctgcctccggccttctaaagtgtctgggattataggca	25950
tgagccactgtgcggcgtctcaaccccttttttttttttttttttttttttt	26000
attagatggaaatatgtcttagtcccagatatccatccattggttt	26050
attacccttcattatcacttcaaaatttttttttttttttttttttttttttt	26100
tttatacaggtaaaatt	26150
ctatgattt	26200
ctctgtcactcaggctggagtgcgggttgcgtatcatggctactatggc	26250
ctcgaccccttctggcgtcaagtgcacccctccatcgtggctccaaagctgag	26300
actacaggcatgcacccacacatctagctaatttttttttttttttttttttt	26350
aacaaggcttt	26400
gggatccctctgtctcagccctacccaaatgtctgggattacaggcatgagc	26450

Fig. 16 (continued)

25/34

Fig. 16 (continued)

26/34

Fig. 16 (continued)

27/34

Fig. 16 (continued)

28/34

tttgtcttgtgtacatgtgtttgttatgtgtgtgtctaaaagtt	34900
tggctttgagcttgccttgaattcttggatgaacaataaccaagaatac	34950
ttaaactctgatcatcttgcacagatatcccctacaggctatggccttt	35000
gaattgtgcctccaggataaaaaaggcagcaagcagcatactgcttcag	35050
attcatgggtgcacatgtgagggtgaaaaaaaaaaaagatgaatccta	35100
ttaaatgcccccagataacagtgataactctttgttagataacttttg	35150
cttgcactggttcattaaataaggacataagtaagatctatTTTGT	35200
ctcttctcccaaccaccacaactagGATTATTGGCTATCTCTGT	35250
D Y W L S L L F	
CAAGAAATTGGTGGGCACCAAGGTGTTATGGCAAGCGTGCAAGGTCAA	35300
K K L V G T K V L M A S V Q G S	
AGAGAAGGAAGCTTCGAGTATACTTCATTGCACAAACACTGACAAGtaa	35350
K . R . R . K . L . R V Y L H C T N . T . D N	
gtatgaaaacacaccccttaccaatcatcaagttttagggtaagcctgt	35400
aactttactcaaacaccccttgcatgtgtctatacatgtcataagtata	35450
ggcagttgcattttatggataaaaatgtacttttttttttttttttttttt	35500
ttttagaaagaaaaatgtacttttttttttttttttttttttttttttttt	35550
ctcgctcgtcacccccaggctggaggcgtggcaatctcagctcactgc	35600
aacctccgcctcccggttcaagtgattcttgaagaggagaacaataata	35650
acaacaatattatTTTcaaaagtgtgaccgcagttctggagttgagaa	35700
gacatcgagatTTTtagcctcatactcttgcttttagtgcataaaaaat	35750
gttcctaaatctcaggaatattcttagataggtttcaatctatcattcc	35800
tgataagatgtgtgaaataactaattctaggccaaaaagaccagctacc	35850
atttccgatt	35900
gaagtagcgaggggaatggttgaatggataattcataaaaaatgtcag	35950
tagatTTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTT	36000
ccccctgttt	36050
gccttttacccctctatgtctgaaaaatggatagtctggctggcctttaa	36100
caacccagctggcagagctgtgaggatctcagttgtcgtctagccagaca	36150
ttggtagcatgaacggcaacatTTTattgttttttttttttttttttttttt	36200
caactcggtctaaaacgatcataaaagaaggatactaagagggccact	36250
gtcattatggatcctaataactttagatgcattatggattgtcattatgga	36300
tactaataacttaggatcacattttaatttttttttttttttttttttttttt	36350
agatacatatttctatTTTcaagttaacctcttgcTTTTTTTTTTTTTTTT	36400
P R Y	
AAGAAGGAGATTTAACCTCTGTATGCCATAAACCTCCATAATGTCACCAAG	36450
K . E . G . D . L . T . L . Y . A . I . N . L . H . N . V . T . K	
TACTTGCAGTTACCTATCCTTTCTAACAAGCAAGTGATAAAATACCT	36500
Y . L . R . L . P . Y . P . F . S . N . K . Q . V . D . K . Y . L	
TCTAAGACCTTGGGACCTCATGGATTACTTCCAAAgtaagtatTTTCC	36550
L . B . R . L . G . P . H . G . L . L . S . K	
ttgttcattccaaacttcaataaattttatttttttttttttttttttttttt	36600
agtttggacagggagcaaaagacaaagtcaactatatactcaagtctataaa	36650
ttcttaatatttcaggaaattttatgttatgttatgttatgttatgttatgt	36700
actcatcctaagagtctaaagcaaaaggatgtgacacacaaactagcagtt	36750
atcttagagataatTTTgcatttttttttttttttttttttttttttttttttt	36800
actcaacgcatt	36850
acatttcactaaagcaaaatataccttttttttttttttttttttttttttttt	36900
tgggttgttataaaatatacatttttttttttttttttttttttttttttttt	36950
acatt	37000
ttcttaagttagattctcatacacttttttttttttttttttttttttttttt	37050
acatcttataaaaggcctgtatttttttttttttttttttttttttttttttt	37100
tcttaaaggatTTTtaggttttttttttttttttttttttttttttttttttt	37150
ttggaaaggatTTTtt	37200
attt	37250
gaagctgttt	37300
ggcaatatggcaaaactctcttttttttttttttttttttttttttttttttt	37350
tgggttgttt	37400
attgttt	37450
ctgttt	37500
taaataaaaattaagatTTTtttttttttttttttttttttttttttttttttt	37550

Fig. 16 (continued)

29/34

Fig. 16 (continued)

30/34

Fig. 16 (continued)

31/34

ctttgtcagcaataatgtgagaggacagattgttagatatgatagtat	43450
aaaaaaatggtaatgacaattcagaggcgaggagattctgtaaactaaa	43500
attactataatgaaattgatttgtcaagaggataaatttagaaaaacac	43550
ccaataccttataactgtctgttaatgctgttttcttaccttctt	43600
ccttggcagttggaaagctttggctgcaagtaacagaaaactcttaat	43650
tcaaatggcttaagcaataaggaaatgtatattcccacataactagacgt	43700
tcaaaacaggccaggctccagcacttcagtgacgtcaccaggatctgggtt	43750
cttcccagctctctgtctgcacatctttagcgctggcttcatctcagac	43800
tctggtagcatgtatggctgttagctgtttcatggcccccctcaaacctcat	43850
agcaaccagaggaagaaaatgagccatttttgagtcctccatcatagact	43900
tgaataacttttcagagcttctcacagacaaaccttcctcatgtctc	43950
ctcatgtcttattgttcagaaatggtaatgtggccatttcaccagtac	44000
tgccaacaacaacgagggttccataattgtctctgagtaaccctttggaa	44050
tggagagggttgggtcagtctacaactgaacactgcagttctgcgtt	44100
tttaccagtgaaaaaatgtatattttccctttaaggattaatattc	44150
ttcaaatgtatgcctgttagatatagtatctttaaaatttttatttt	44200
aatacgctttagggtagacacacttttgcattacaggggtaattgttagt	44250
ggtgaagactcggctttatgtacttgtcacctgagtgatgtacattgt	44300
acccaataggtatattttcatccattaccctccgcgccttccctt	44350
ctgagtcctcaacatccctataccactgtgtatgttcttgttac	44400
agctaagcttccacttataagtgagaacatgcagtattggtttccatt	44450
cctgagttacttcccttaggataacagccccagttccgtccaagttgct	44500
gcaaaatacattattcttctttatggctgagtaatagtccatggtacata	44550
tataccacattttcttccattacttgcgttgcacacttaggttac	44600
ttccattcaatttcatcaatttaagtatattgttagggactaaagctg	44650
aaaattaaatttttagatcttcaactactttaattttatgttagtgg	44700
tttttatattttcacattttgaaataaaagtaattttataacccattgtat	44750
gtatgactattcttttagtaatgtaaaggcttacagactccatattgga	44800
accactagtgtgttgcacccctgttatactatcaggatcctcga	44898

Figure 17

human	MLLRSKPALP PPLMLLLLGP LGPLSPGALP RPAQAOQDVVD LDFFTQEPLH	50
mouse	~~~~~ML RLLLWLWGP LGALAQGAPA GTAPTDVVVD LEFYTKRPLR	
rat	~~~~~ ~LLLWLWGR LRALTOGTPA GTAPTKDVVD LEFYTKRLFQ	
human	LVSPSFLSVT IDANLATDPR FLILLGSPKL RTLARGLSPA YLRFGGKTD	100
mouse	SVSPSFLSIT IDASLATDPR FLTFLGSPPRL RALARGLSPA YLRFGGKTD	
rat	SVSPSFLSIT IDASLATDPR FLTFLSSPPRL RALSRLGLSPA YLRFGGKTD	
human	FLIFDPKES TFEERSYWQS QVNQDICKYQ SIPPDVEEKL RLEWPIQEQL	150
mouse	FLIFDPDKEP TSEERSYWKS QVNHDICRSE PVSAAVLRLQ QVEWPFQELL	
rat	FLIFDPNNEP TSEERSYWQS QDNNDICGSD RVSADVL~~~	
human	LLREHYQKKF KNSTYSRSSV DVLYTFANCS GLDLIFGLNA LLRTADLWN	200
mouse	LLREQYQKEF KNSTYSRSSV DMLYSFAKCS GLDLIFGLNA LLRTPDLRWN	
rat	~~~~~ ~~~~~ ~~~~~ ~~~~~ ~~~~~ ~~~~~ ~~~~~	
human	SSNAQLLLDY CSSKGYNISW ELGNEPNSFL KKADIFINGS QLGEDYIQLH	250
mouse	SSNAQLLLDY CSSKGYNISW ELGNEPNSFW KKAHILIDGL QLGEDFVELH	
rat	~~~~~ ~~~~~ ~~~~~ ~~~~~ ~~~~~ ~~~~~ ~~~~~	
human	KLLRKSTFKN AKLYGPDVGQ PRRKTAKMLK SFLKAGGEVI DSVTWHHYYL	300
mouse	KLLQSAFQN AKLYGPDIGQ PRGKTVKLLR SFLKAGGEVI DSLTWHHYYL	
rat	~~~~~ ~~~~~ ~~~~~ ~~~~~ ~~~~~ ~~~~~ ~~~~~	
human	NGRTATREDF INPDVLDIFI SSVQKVFQVV ESTRPGKKVW LGETSSAYGG	350
mouse	NGRIATKEDF LSSDALDTFI LSVQKILKVT KEITPGKKVW LGETSSAYGG	
rat	~~~~~ ~~~~~ ~~~~~ ~~~~~ ~~~~~ ~~~~~ ~~~~~	
human	GAPLLSDTFA AGFMWLDKLG LSARMGIEVV MRQVFFGAGN YHLVDENFDP	400
mouse	GAPLLSNTFA AGFMWLDKLG LSAQMGIEVV MRQVFFGAGN YHLVDENFEP	
rat	~~~~~ ~~~~~ ~~~~~ ~~~~~ ~~~~~ ~~~~~ ~~~~~	
human	LPDYWLSLLF KKLVGTVLMS ASVQGSKRRK LRVYLHCTNT DNPRYKEGDL	450
mouse	LPDYWLSLLF KKLVGPRVLL SRVKGPDRSK LRVYLHCTNV YHPRYQEGDL	
rat	~~~~~ ~~~~~ ~~~~~ ~~~~~ ~~~~~ ~~~~~ ~~~~~	
human	TLYAINLHNV TKYLRLPYPF SNKQVDKYLL RPLGPHGLLS KSVQLNGLTL	500
mouse	TLYVNLHNV TKHLKVPPPL FRKPVDTYLL KPSGPDGLLS KSVQLNGQIL	
rat	~~~~~ ~~~~~ ~~~~~ ~~~~~ ~~~~~ ~~~~~ ~~~~~	
human	KMVDDQTLPP LMEKPLRPAGS SLGLPAFSYS FFVIRNAKVA ACI~	543
mouse	KMVDEQTLPA LTEKPLPAGS ALSLPAFSYG FFVIRNAKIA ACI~	
rat	KMVDEQTXPA LTEKPLPAGS SLSVPAFSYG FFVIRNAKIA ACI~	

33/34

Figure 18

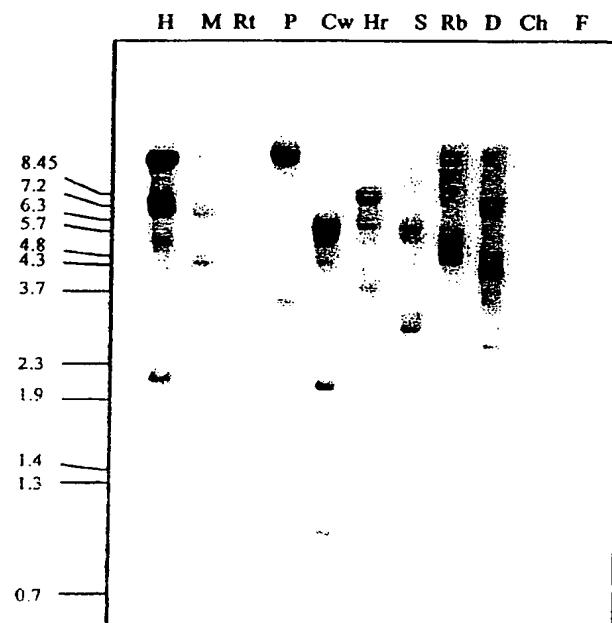


Figure 19

1

SEQUENCE LISTING

(1) GENERAL INFORMATION:

(i) APPLICANT: Iris Pecker, Israel Vlodavsky and Elena Feinstein

(ii) TITLE OF INVENTION: POLYNUCLEOTIDE ENCODING A POLYPEPTIDE HAVING HEPARANASE ACTIVITY AND EXPRESSION OF SAME IN GENETICALLY MODIFIED CELLS

(iii) NUMBER OF SEQUENCES: 47

(iv) CORRESPONDENCE ADDRESS:

(A) ADDRESSEE: Mark M. Friedman c/o Anthony Castorina

(B) STREET: 2001 Jefferson Davis Highway, Suite 207

(C) CITY: Arlington

(D) STATE: Virginia

(E) COUNTRY: United States of America

(F) ZIP: 22202

(v) COMPUTER READABLE FORM:

(A) MEDIUM TYPE: 1.44 megabyte, 3.5" microdisk

(B) COMPUTER: Twinhead* Slimnote-890TX

(C) OPERATING SYSTEM: MS DOS version 6.2, Windows version 3.11

(D) SOFTWARE: Word for Windows version 2.0 converted to an ASCII file

(vi) CURRENT APPLICATION DATA:

(A) APPLICATION NUMBER:

(B) FILING DATE:

(C) CLASSIFICATION:

(vii) PRIOR APPLICATION DATA:

(A) APPLICATION NUMBER: 08/922,170

(B) FILING DATE: 2 SEP 1997

(A) APPLICATION NUMBER: 09/109,386

(B) FILING DATE: 10 JUL 1998

(A) APPLICATION NUMBER: PCT/US98/17954

(B) FILING DATE: 31 AUG 1998

(A) APPLICATION NUMBER: 09/258,892

(B) FILING DATE: 1 MAR 1999

(viii) ATTORNEY/AGENT INFORMATION:

(A) NAME: Friedmam, Mark M.

(B) REGISTRATION NUMBER: 33,883

(C) REFERENCE/DOCKET NUMBER: 910/62

(ix) TELECOMMUNICATION INFORMATION:

(A) TELEPHONE: 972-3-5625553

(B) TELEFAX: 972-3-5625554

(C) TELEX:

(2) INFORMATION FOR SEQ ID NO:1:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 27

(B) TYPE: nucleic acid

(C) STRANDEDNESS: single

(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:1:
CCATCTTAAT ACGACTCACT ATAGGGC 27

(2) INFORMATION FOR SEQ ID NO:2:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 24

(B) TYPE: nucleic acid

(C) STRANDEDNESS: single

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:2:
GTAGTGATGC CATGTAAGT AATC 24

(2) INFORMATION FOR SEQ ID NO:3:
(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 23
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear
(xi) SEQUENCE DESCRIPTION: SEQ ID NO:3:
ACTCACTATA GGGCTCGAGC GGC 23

(2) INFORMATION FOR SEQ ID NO:4:
(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 22
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear
(xi) SEQUENCE DESCRIPTION: SEQ ID NO:4:
GCATCTTAGC CGTCTTCTT CG 22

(2) INFORMATION FOR SEQ ID NO:5:
(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 15
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear
(xi) SEQUENCE DESCRIPTION: SEQ ID NO:5:
TTTTTTTTT TTTTT 15

(2) INFORMATION FOR SEQ ID NO:6:
(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 23
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear
(xi) SEQUENCE DESCRIPTION: SEQ ID NO:6:
TTCGATCCC AGAAGGAATC AAC 23

(2) INFORMATION FOR SEQ ID NO:7:
(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 24
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear
(xi) SEQUENCE DESCRIPTION: SEQ ID NO:7:
GTAGTGATGC CATGTAAGT AATC 24

(2) INFORMATION FOR SEQ ID NO:8:
(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 9
(B) TYPE: amino acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear
(xi) SEQUENCE DESCRIPTION: SEQ ID NO:8:
Tyr Gly Pro Asp Val Gly Gln Pro Arg

(2) INFORMATION FOR SEQ ID NO:9:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 1721
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: double
- (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:9:

CTAGAGCTTT CGACTCTCCG CTGCGCGCA GCTGGCGGGG GGAGCAGCCA GGTGAGCCCA 60
 AGATGCTGCT GCGCTCGAAG CCTCGCTGC CGCCGCCGCT GATGCTGCTG CTCCTGGGC 120
 CGCTGGTCC CCTCTCCCCCT GGCGCCCTGC CCCGACCTGC GCAAGCAGAG GACGTCGTGG 180
 ACCTGGACTT cTTCACCCAG GAGCGCTGC ACCTGGTGAG CCCCTCGTTC CTGTCGTCA 240
 CCATTGACGC CAACCTGGCC ACGGACCCGC GGTTCCAT CCTCCTGGGT TCTCAAAGC 300
 TTCTGACCTT GGCCAGAGGC TTGCTCTCTG CGTACCTGAG GTTGTGGC ACCAAGACAG 360
 ACTTCCTAAT TTTCGATCCC AAGAAGGAAT CAACCTTGAG AGAGAGAAAGT TACTGGCAAT 420
 CTCAAGTCAA CCAGGATATT TGCAAAATATG GATCCATCCC TCTGATGTG GAGGAGAAAGT 480
 TACGGTTGGA ATGGCCCTAC CAGGAGCAAT TGCTACTCCG AGAACACTAC CAGAAAAAGT 540
 TCAAGAACAG CACCTACTCA AGAACCTCTG TAGATGTGCT ATACACTTTT GCAAACGTGCT 600
 CAGGACTGGA TTGATCTT GGCTAAATG CGTTATTAAG AACAGCAGAT TTGCACTGG 660
 ACAGTTCTAA TGCTCAGTTG CTCCTGGACT ACTGCTCTTC CAAGGGGTAT AACATTTCTT 720
 GGGAACTAGG CAATGAACCT AACAGTTTCC TTAAGAAGGC TGATATTTTC ATCAATGGGT 780
 CGCACTTGG AGAAAGATTAT ATTCAATTGC ATAAACTTCT AAGAAAGTCC ACCTTCAAAAA 840
 ATGAAAACACT CTATGGTCT GATGTTGGTC AGCCTCGAAG AAAGACGGCT AAGATGCTGA 900
 AGAGCTTCCT GAAGGCTGGT GGAGAAGTGA TTGATTCACTG TACATGGCAT CACTACTATT 960
 TGAATGGACG GACTGCTACC AGGGAAAGATT TTCTAAACCC TGATGTATTG GACATTTTA 1020
 TTTCATCTGT GCAAAAGTT TTCCAGGGG TTGAGAGCAC CAGGCTGGC AAGAAGGTCT 1080
 GGTTAGGAGA AACAAGCTCT GCATATGGAG GCGGAGCGCC CTTGCTATCC GACACCTTTG 1140
 CAGCTGGCTT TATGTGGCTG GATAAAATGG GCCTGTCAGC CCGAATGGGA ATAGAAGTGG 1200
 TGATGAGGCA AGTATTCTT GGAGCAGGAA ACTACCATTT AGTGGATGAA AACTTCGATC 1260
 CTTTACCTGA TTATTGGCTA TCTCTCTGT TCAAGAAATT GGTGGGCACC AAGGTGTTAA 1320
 TGGCAAGCGT GCAAGGTTCA AAGAGAAGGA AGCTTCGAGT ATACCTTCAT TGCAACAAACA 1380
 CTGACAATCC AAGGTATAAA GAAGGAGATT TAACTCTGTA TGCCATAAAC CTCCATAACG 1440
 TCACCAAGTA CTTGCGGTTA CCCTATCCTT TTTCTAACAA GCAAGTGGAT AAATACCTTC 1500
 TAAGACCTTT GGGACCTCAT GGATTACTTT CCAAATCTGT CCAACTCAAT GGTCTAACTC 1560
 TAAAGATGGT GGATGATCAA ACCTTGGCAC CTTTAATGGA AAAACCTCTC CGGCCAGGAA 1620
 GTTCACTGGG CTTGCCAGCT TTCTCATATA GTTTTTTGT GATAAGAAAT GCCAAAGTTG 1680
 CTGCTTGCAT CTGAAAATAA AATATACTAG TCCTGACACT G 1721

(2) INFORMATION FOR SEQ ID NO:10:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 543
- (B) TYPE: amino acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:10:

Met Leu Leu Arg Ser Lys Pro Ala Leu Pro Pro Pro Leu Met Leu Leu

Leu Leu Gly Pro Leu Gly Pro Leu Ser Pro Gly Ala Leu Pro Arg Pro

Ala Gln Ala Gln Asp Val Val Asp Leu Asp Phe Phe Thr Gln Glu Pro

Leu His Leu Val Ser Pro Ser Phe Leu Ser Val Thr Ile Asp Ala Asn

Leu Ala Thr Asp Pro Arg Phe Leu Ile Leu Leu Gly Ser Pro Lys Leu
65 70 75 80

Arg Thr Leu Ala Arg Gly Leu Ser Pro Ala Tyr Leu Arg Phe Gly Gly
85 90 95

Thr Lys Thr Asp Phe Leu Ile Phe Asp Pro Lys Lys Glu Ser Thr Phe
100 105 110

Glu Glu Arg Ser Tyr Trp Gln Ser Gln Val Asn Gln Asp Ile Cys Lys
115 120 125

Tyr Gly Ser Ile Pro Pro Asp Val Glu Glu Lys Leu Arg Leu Glu Trp
130 135 140

Pro Tyr Gln Glu Gln Leu Leu Leu Arg Glu His Tyr Gln Lys Lys Phe
145 150 155 160

Lys Asn Ser Thr Tyr Ser Arg Ser Ser Val Asp Val Leu Tyr Thr Phe
165 170 175

Ala Asn Cys Ser Gly Leu Asp Leu Ile Phe Gly Leu Asn Ala Leu Leu
180 185 190

Arg Thr Ala Asp Leu Gln Trp Asn Ser Ser Asn Ala Gln Leu Leu Leu
195 200 205

Asp Tyr Cys Ser Ser Lys Gly Tyr Asn Ile Ser Trp Glu Leu Gly Asn
210 215 220

Glu Pro Asn Ser Phe Leu Lys Lys Ala Asp Ile Phe Ile Asn Gly Ser
225 230 235 240

Gln Leu Gly Glu Asp Tyr Ile Gln Leu His Lys Leu Leu Arg Lys Ser
245 250 255

Thr Phe Lys Asn Ala Lys Leu Tyr Gly Pro Asp Val Gly Gln Pro Arg
260 265 270

Arg Lys Thr Ala Lys Met Leu Lys Ser Phe Leu Lys Ala Gly Gly Glu
275 280 285

Val Ile Asp Ser Val Thr Trp His His Tyr Tyr Leu Asn Gly Arg Thr
290 295 300

Ala Thr Arg Glu Asp Phe Leu Asn Pro Asp Val Leu Asp Ile Phe Ile
305 310 315 320

Ser Ser Val Gln Lys Val Phe Gln Val Val Glu Ser Thr Arg Pro Gly
325 330 335

Lys Lys Val Trp Leu Gly Glu Thr Ser Ser Ala Tyr Gly Gly Ala
340 345 350

Pro Leu Leu Ser Asp Thr Phe Ala Ala Gly Phe Met Trp Leu Asp Lys
355 360 365

Leu Gly Leu Ser Ala Arg Met Gly Ile Glu Val Val Met Arg Gln Val
 370 375 380

Phe Phe Gly Ala Gly Asn Tyr His Leu Val Asp Glu Asn Phe Asp Pro
 385 390 395 400

Leu Pro Asp Tyr Trp Leu Ser Leu Leu Phe Lys Lys Leu Val Gly Thr
 405 410 415

Lys Val Leu Met Ala Ser Val Gln Gly Ser Lys Arg Arg Lys Leu Arg
 420 425 430

Val Tyr Leu His Cys Thr Asn Thr Asp Asn Pro Arg Tyr Lys Glu Gly
 435 440 445

Asp Leu Thr Leu Tyr Ala Ile Asn Leu His Asn Val Thr Lys Tyr Leu
 450 455 460

Arg Leu Pro Tyr Pro Phe Ser Asn Lys Gln Val Asp Lys Tyr Leu Leu
 465 470 475 480

Arg Pro Leu Gly Pro His Gly Leu Leu Ser Lys Ser Val Gln Leu Asn
 485 490 495

Gly Leu Thr Leu Lys Met Val Asp Asp Gln Thr Leu Pro Pro Leu Met
 500 505 510

Glu Lys Pro Leu Arg Pro Gly Ser Ser Leu Gly Leu Pro Ala Phe Ser
 515 520 525

Tyr Ser Phe Phe Val Ile Arg Asn Ala Lys Val Ala Ala Cys Ile
 530 535 540 543

(2) INFORMATION FOR SEQ ID NO:11:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 1721
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: double
- (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:11:

CT AGA GCT TTC GAC 14

TCT CCG CTG CGC GGC AGC TGG CCG GGG GAG CAG CCA GGT GAG CCC AAG 62

ATG CTG CTG CGC TCG AAG CCT GCG CTG CCG CCG CCG CTG ATG CTG CTG 110
 Met Leu Leu Arg Ser Lys Pro Ala Leu Pro Pro Leu Met Leu Leu
 5 10 15

CTC CTG GGG CCG CTG GGT CCC CTC TCC CCT GGC GCC CTG CCC CGA CCT 158
 Leu Leu Gly Pro Leu Gly Pro Leu Ser Pro Gly Ala Leu Pro Arg Pro
 20 25 30

GCG CAA GCA CAG GAC GTC GTG GAC CTG GAC TTC TTC ACC CAG GAG CCG 206
 Ala Gln Ala Gln Asp Val Val Asp Leu Asp Phe Phe Thr Gln Glu Pro
 35 40 45

CTG CAC CTG GTG AGC CCC TCG TTC CTG TCC GTC ACC ATT GAC GCC AAC 254
 Leu His Leu Val Ser Pro Ser Phe Leu Ser Val Thr Ile Asp Ala Asn
 50 55 60

CTG GCC ACG GAC CCG CGG TTC CTC ATC CTC CTG GGT TCT CCA AAG CTT 302
 Leu Ala Thr Asp Pro Arg Phe Leu Ile Leu Leu Gly Ser Pro Lys Leu
 65 70 75 80

CGT ACC TTG GCC AGA GGC TTG TCT CCT GCG TAC CTG AGG TTT GGT GGC 350
 Arg Thr Leu Ala Arg Gly Leu Ser Pro Ala Tyr Leu Arg Phe Gly Gly
 85 90 95

ACC AAG ACA GAC TTC CTA ATT TTC GAT CCC AAG AAG GAA TCA ACC TTT 398
 Thr Lys Thr Asp Phe Leu Ile Phe Asp Pro Lys Lys Glu Ser Thr Phe
 100 105 110

GAA GAG AGA AGT TAC TTG CAA TCT CAA GTC AAC CAG GAT ATT TGC AAA 446
 Glu Glu Arg Ser Tyr Trp Gln Ser Gln Val Asn Gln Asp Ile Cys Lys
 115 120 125

TAT GGA TCC ATC CCT CCT GAT GTG GAG AAG TTA CGG TTG GAA TGG 494
 Tyr Gly Ser Ile Pro Pro Asp Val Glu Glu Lys Leu Arg Leu Glu Trp
 130 135 140

CCC TAC CAG GAG CAA TTG CTA CTC CGA GAA CAC TAC CAG AAA AAG TTC 542
 Pro Tyr Gln Glu Gln Leu Leu Leu Arg Glu His Tyr Gln Lys Lys Phe
 145 150 155 160

AAG AAC AGC ACC TAC TCA AGA AGC TCT GTA GAT GTG CTA TAC ACT TTT 590
 Lys Asn Ser Thr Tyr Ser Arg Ser Ser Val Asp Val Leu Tyr Thr Phe
 165 170 175

GCA AAC TGC TCA GGA CTG GAC TTG ATC TTT GGC CTA AAT GCG TTA TTA 638
 Ala Asn Cys Ser Gly Leu Asp Leu Ile Phe Gly Leu Asn Ala Leu Leu
 180 185 190

AGA ACA GCA GAT TTG CAG TGG AAC AGT TCT AAT GCT CAG TTG CTC CTG 686
 Arg Thr Ala Asp Leu Gln Trp Asn Ser Ser Asn Ala Gln Leu Leu
 195 200 205

GAC TAC TGC TCT TCC AAG GGG TAT AAC ATT TCT TGG GAA CTA GGC AAT 734
 Asp Tyr Cys Ser Ser Lys Gly Tyr Asn Ile Ser Trp Glu Leu Gly Asn
 210 215 220

GAA CCT AAC AGT TTC CTT AAG AAG GCT GAT ATT TTC ATC AAT GGG TCG 782
 Glu Pro Asn Ser Phe Leu Lys Lys Ala Asp Ile Phe Ile Asn Gly Ser
 225 230 235 240

CAG TTA GGA GAA GAT TAT ATT CAA TTG CAT AAA CTT CTA AGA AAG TCC 830
 Gln Leu Gly Asp Tyr Ile Gln Leu His Lys Leu Leu Arg Lys Ser
 245 250 255

ACC TTC AAA AAT GCA AAA CTC TAT GGT CCT GAT GTT GGT CAG CCT CGA 878
 Thr Phe Lys Asn Ala Lys Leu Tyr Gly Pro Asp Val Gly Gln Pro Arg
 260 265 270

7

AGA AAG ACG GCT AAG ATG CTG AAG AGC TTC CTG AAG GCT GGT GGA GAA 926
 Arg Lys Thr Ala Lys Met Leu Lys Ser Phe Leu Lys Ala Gly Gly Glu
 275 280 285

GTG ATT GAT TCA GTT ACA TGG CAT CAC TAC TAT TTG AAT GGA CGG ACT 974
 Val Ile Asp Ser Val Thr Trp His His Tyr Tyr Leu Asn Gly Arg Thr
 290 295 300

GCT ACC AGG GAA GAT TTT CTA AAC CCT GAT GTA TTG GAC ATT TTT ATT 1022
 Ala Thr Arg Glu Asp Phe Leu Asn Pro Asp Val Leu Asp Ile Phe Ile
 305 310 315 320

TCA TCT GTG CAA AAA GTT TTC CAG GTG GTT GAG AGC ACC AGG CCT GGC 1070
 Ser Ser Val Gln Lys Val Phe Gln Val Val Glu Ser Thr Arg Pro Gly
 325 330 335

AAG AAG GTC TGG TTA GGA GAA ACA AGC TCT GCA TAT GGA GGC GGA GCG 1118
 Lys Lys Val Trp Leu Gly Glu Thr Ser Ser Ala Tyr Gly Gly Ala
 340 345 350

CCC TTG CTA TCC GAC ACC TTT GCA GCT GGC TTT ATG TGG CTG GAT AAA 1166
 Pro Leu Leu Ser Asp Thr Phe Ala Ala Gly Phe Met Trp Leu Asp Lys
 355 360 365

TTG GGC CTG TCA GCC CGA ATG GGA ATA GAA GTG GTG ATG AGG CAA GTA 1214
 Leu Gly Leu Ser Ala Arg Met Gly Ile Glu Val Val Met Arg Gln Val
 370 375 380

TTC TTT GGA GCA GGA AAC TAC CAT TTA GTG GAT GAA AAC TTC GAT CCT 1262
 Phe Phe Gly Ala Gly Asn Tyr His Leu Val Asp Glu Asn Phe Asp Pro
 385 390 395 400

TTA CCT GAT TAT TGG CTA TCT CTT CTG TTC AAG AAA TTG GTG GGC ACC 1310
 Leu Pro Asp Tyr Trp Leu Ser Leu Leu Phe Lys Lys Leu Val Gly Thr
 405 410 415

AAG GTG TTA ATG GCA AGC GTG CAA GGT TCA AAG AGA AGG AAG CTT CGA 1358
 Lys Val Leu Met Ala Ser Val Gln Gly Ser Lys Arg Arg Lys Leu Arg
 420 425 430

GTA TAC CTT CAT TGC ACA AAC ACT GAC AAT CCA AGG TAT AAA GAA GGA 1406
 Val Tyr Leu His Cys Thr Asn Thr Asp Asn Pro Arg Tyr Lys Glu Gly
 435 440 445

GAT TTA ACT CTG TAT GCC ATA AAC CTC CAT AAC GTC ACC AAG TAC TTG 1454
 Asp Leu Thr Leu Tyr Ala Ile Asn Leu His Asn Val Thr Lys Tyr Leu
 450 455 460

CGG TTA CCC TAT CCT TTT TCT AAC AAG CAA GTG GAT AAA TAC CTT CTA 1502
 Arg Leu Pro Tyr Pro Phe Ser Asn Lys Gln Val Asp Lys Tyr Leu Leu
 465 470 475 480

AGA CCT TTG GGA CCT CAT GGA TTA CTT TCC AAA TCT GTC CAA CTC AAT 1550
 Arg Pro Leu Gly Pro His Gly Leu Leu Ser Lys Ser Val Gln Leu Asn
 485 490 495

GGT CTA ACT CTA AAG ATG GTG GAT GAT CAA ACC TTG CCA CCT TTA ATG 1598

Gly Leu Thr Leu Lys Met Val Asp Asp Gln Thr Leu Pro Pro Leu Met
 500 505 510

GAA AAA CCT CTC CGG CCA GGA AGT TCA CTG GGC TTG CCA GCT TTC TCA 1646
 Glu Lys Pro Leu Arg Pro Gly Ser Ser Leu Gly Leu Pro Ala Phe Ser
 515 520 525

TAT AGT TTT TTT GTG ATA AGA AAT GCC AAA GTT GCT GCT TGC ATC TGA 1694
 Tyr Ser Phe Phe Val Ile Arg Asn Ala Lys Val Ala Ala Cys Ile
 530 535 540 543

AAA TAA AAT ATA CTA GTC CTG ACA CTG 1721

(2) INFORMATION FOR SEQ ID NO:12:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 824
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: double
- (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:12

CTGGCAAGAA GGTCTGGTTG GGAGAGACGA GCTCAGCTTA CGGTGGCGGT GCACCCCTTGC 60
 TGTCCAACAC CTTTGCAGCT GGCTTTATGT GGCTGGATAA ATTGGGCTG TCAGCCCAGA 120
 TGGGCATAGA AGTCGTGATG AGGCAGGTGT TCTTCGGAGC AGGCAACTAC CACTTAGTGG 180
 ATGAAAACCTT TGAGCCTTAA CCTGATTACT GGCTCTCTCT TCTGTTCAAG AAACCTGGTAG 240
 GTCCCAGGGT GTTACTGTCA AGAGTGAAGAAG GCCCAGACAG GAGCAAACTC CGAGTGTATC 300
 TCCACTGCACTAACGCTAT CACCCACGAT ATCAGGAAGG AGATCTAACT CTGTATGTCC 360
 TGAACCTCCA TAATGTCACTTAAAGCAGCTTGA AGGTACCGCC TCCGTTGTTG AGGAAACCCAG 420
 TGGATACGTA CCTTCTGAAG CCTTCGGGGC CGGATGGATT ACTTTCCAAA TCTGTCCAAC 480
 TGAACCGGTCA AATTCTGAAG ATGGTGGATG AGCAGACCCCT GCCAGCTTTG ACAGAAAAAC 540
 CTCCTCCCCGC AGGAAGTGCA CTAACCTGC CTGCCTTTTC CTATGGTTT TTGTCTAA 600
 GAAATGCCAA ATACGCTGCT TGTATATGAA AATAAAAGGC ATACGGTACC CCTGAGACAA 660
 AAGCCGAGGG GGGTGTATT CATAAAACAA AACCCCTAGTT TAGGAGGCCA CCTCCTTGCC 720
 GAGTTCCAGA GCTTCGGGAG GGTGGGGTAC ACTTCAGTAT TACATTCACT GTGGTGTCT 780
 CTCTAAGAAG AATACTGCAG GTGGTGACAG TTAATAGCAC TGTG 824

(2) INFORMATION FOR SEQ ID NO:13:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 1899
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: double
- (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:13

GGGAAAGCGA GCAAGGAAGT AGGAGAGAGC CGGGCAGGCG GGGCGGGGTT GGATTGGGAG 60
 CAGTGGGAGG GATCGAGAAG AGGAGTGGGA GGGATGGAGG GCGCAGTGGG AGGGGTGAGG 120
 AGGCGTAACG GGGCGGAGGA AAGGAGAAA GGGCGCTGGG GCTCGGCGGG AGGAAGTGC 180
 AGAGCTCTCG ACTCTCCGCT GCGCGGAGC TGGCGGGGG AGCAGGCCAGG TGAGCCAA 240
 ATGCTGCTGC GCTCGAAGCC TGCCTGCGC CGGCCGCTGA TGCTGCTGCT CCTGGGGCCG 300
 CTGGGTCCCC TCTCCCTGG CGCCCTGCC CGACCTGCGC AAGCACAGGA CGTCGTGGAC 360
 CTGGACTTCT TCACCCAGGA GCCGCTGCAC CTGGTGAGCC CCTCGTTCCCT GTCCGTCACC 420
 ATTGACGCCA ACCTGGCCAC GGACCCGCGG TTCCCTCATCC TCCTGGGGTTC TCCAAAGCTT 480
 CGTACCTTGG CCAGAGGCTT GTCTCCTGCG TACCTGAGGT TTGGTGGCAC CAAGACAGAC 540
 TTCCTAATTT TCGATCCCAA GAAGGAATCA ACCTTTGAAG AGAGAAGTTA CTGGCAATCT 600
 CAAGTCACC AGGATATTTG CAAATATGGA TCCATCCCTC CTGATGTGGA GGAGAAGTTA 660
 CGGTTGGAAT GGCCCTACCA GGAGCAATTG CTACTCCGAG AACACTACCA GAAAAAGTTC 720
 AAGAACAGCA CCTACTCAAG AAGCTCTGTA GATGTGCTAT ACACCTTTGC AACTGCTCA 780

GGACTGGACT	TGATCTTGG	CCTAAATGCG	TTATTAAGAA	CAGCAGATTT	GCAGTGGAAC	840
AGTTCTAATG	CTCAGTTGCT	CCTGGACTAC	TGCTCTTCCA	AGGGTATAA	CATTTCTTGG	900
GAACCTAGGCA	ATGAACCTAA	CAGTTCCCTT	AAGAAGGCTG	ATATTTCAT	CAATGGGTG	960
CAGTTAGGAG	AAGATTATAT	TCAATTGCGAT	AAACTCTAA	GAAAGTCCAC	CTTCAAAAAT	1020
GCAAAACTCT	ATGGTCCCTGA	TGTTGGTCAG	CCTCGAAGAA	AGACGGCTAA	GATGCTGAAG	1080
AGCTTCCCTGA	AGGCTGGTGG	AGAAGTGATT	GATTCACTTA	CATGGCATCA	CTACTATTTG	1140
AATGGACGGA	CTGCTACCAAG	GGAAAGATTTC	CTAACCCCTG	ATGTATTGGA	CATTTTTATT	1200
TCATCTGTGC	AAAAAGTTT	CCAGGTGGTT	GAGAGCACCA	GGCCTGGCAA	GAAGGTCTGG	1260
TTAGGAGAAA	CAAGCTCTGC	ATATGGAGGC	GGAGCGCCCT	TGCTATCCGA	CACCTTGCA	1320
GCTGGCTTTA	TGTGGCTGGA	AAATTGGGC	CTGTCAGCCC	GAATGGGAAT	AGAAGTGGTG	1380
ATGAGGCAAG	TATTCTTGG	AGCAGGAAAC	TACCAATTAG	TGGATGAAAA	CTTCGATCCT	1440
TTACCTGATT	ATTGGCTATC	TCTTCTGTT	AAGAAATTGG	TGGGCACCAA	GGTGTAAATG	1500
GCAAGCGTGC	AAGGTTCAA	GAGAAGGAAG	CTTCGAGTAT	ACCTTCATTG	CACAAACACT	1560
GACAATCAA	GGTATAAAGA	AGGAGATTAA	ACTCTGTATG	CCATAAACCT	CCATAACGTC	1620
ACCAAGTACT	TGCGTTTAC	CTATCCTTT	TCTAACAAAGC	AACTGGATAA	ATACCTTCTA	1680
AGACCTTGG	GACCTCATGG	ATTACTTCC	AAATCTGTCC	AACTCAATGG	TCTAACTCTA	1740
AAGATGGTGG	ATGATCAAAC	CTTGCCACCT	TTAATGGAAA	AACTCTCCG	GCCAGGAAGT	1800
TCACTGGGCT	TGCCAGCTTT	CTCATATAGT	TTTTTGTGA	TAAGAAATGC	CAAAGTTGCT	1860
GCTTGATCT	GAAAATAAAA	TATACTAGTC	CTGACACTG			1899

(2) INFORMATION FOR SEQ ID NO:14:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 592
- (B) TYPE: amino acid
- (C) STRANDEDNESS: singl
- (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:14

Met	Glu	Gly	Ala	Val	Gly	Gly	Val	Arg	Arg	Arg	Asn	Gly	Ala	Glu
5							10							15
Glu	Arg	Arg	Lys	Gly	Arg	Trp	Gly	Ser	Ala	Gly	Gly	Ser	Ala	Arg
20								25						30
Ala	Leu	Asp	Ser	Pro	Leu	Arg	Gly	Ser	Trp	Arg	Gly	Glu	Gln	Pro
35									40					45
Gly	Glu	Pro	Lys	Met	Leu	Leu	Arg	Ser	Lys	Pro	Ala	Leu	Pro	Pro
50									55					60
Pro	Leu	Met	Leu	Leu	Leu	Gly	Pro	Leu	Gly	Pro	Leu	Ser	Pro	
65								70						75
Gly	Ala	Leu	Pro	Arg	Pro	Ala	Gln	Ala	Gln	Asp	Val	Val	Asp	Leu
80								85						90
Asp	Phe	Phe	Thr	Gln	Glu	Pro	Leu	His	Leu	Val	Ser	Pro	Ser	Phe
95								100						105
Leu	Ser	Val	Thr	Ile	Asp	Ala	Asn	Leu	Ala	Thr	Asp	Pro	Arg	Phe
110								115						120
Leu	Ile	Leu	Leu	Gly	Ser	Pro	Lys	Leu	Arg	Thr	Leu	Ala	Arg	Gly
125								130						135
Leu	Ser	Pro	Ala	Tyr	Leu	Arg	Phe	Gly	Gly	Thr	Lys	Thr	Asp	Phe
140								145						150
Leu	Ile	Phe	Asp	Pro	Lys	Lys	Glu	Ser	Thr	Phe	Glu	Glu	Arg	Ser
155								160						165
Tyr	Trp	Gln	Ser	Gln	Val	Asn	Gln	Asp	Ile	Cys	Lys	Tyr	Gly	Ser
170								175						180
Ile	Pro	Pro	Asp	Val	Glu	Glu	Lys	Leu	Arg	Leu	Glu	Trp	Pro	Tyr
185								190						195
Gln	Glu	Gln	Leu	Leu	Leu	Arg	Glu	His	Tyr	Gln	Lys	Lys	Phe	Lys
200								205						210
Asn	Ser	Thr	Tyr	Ser	Arg	Ser	Ser	Val	Asp	Val	Leu	Tyr	Thr	Phe

10

215	220	225
Ala Asn Cys Ser Gly Leu Asp Leu Ile Phe Gly Leu Asn Ala Leu		
230	235	240
Leu Arg Thr Ala Asp Leu Gln Trp Asn Ser Ser Asn Ala Gln Leu		
245	250	255
Leu Leu Asp Tyr Cys Ser Ser Lys Gly Tyr Asn Ile Ser Trp Glu		
260	265	270
Leu Gly Asn Glu Pro Asn Ser Phe Leu Lys Lys Ala Asp Ile Phe		
275	280	285
Ile Asn Gly Ser Gln Leu Gly Glu Asp Tyr Ile Gln Leu His Lys		
290	295	300
Leu Leu Arg Lys Ser Thr Phe Lys Asn Ala Lys Leu Tyr Gly Pro		
305	310	315
Asp Val Gly Gln Pro Arg Arg Lys Thr Ala Lys Met Leu Lys Ser		
320	325	330
Phe Leu Lys Ala Gly Gly Glu Val Ile Asp Ser Val Thr Trp His		
335	340	345
His Tyr Tyr Leu Asn Gly Arg Thr Ala Thr Arg Glu Asp Phe Leu		
350	355	360
Asn Pro Asp Val Leu Asp Ile Phe Ile Ser Ser Val Gln Lys Val		
365	370	375
Phe Gln Val Val Glu Ser Thr Arg Pro Gly Lys Lys Val Trp Leu		
380	385	390
Gly Glu Thr Ser Ser Ala Tyr Gly Gly Ala Pro Leu Leu Ser		
395	400	405
Asp Thr Phe Ala Ala Gly Phe Met Trp Leu Asp Lys Leu Gly Leu		
410	415	420
Ser Ala Arg Met Gly Ile Glu Val Val Met Arg Gln Val Phe Phe		
425	430	435
Gly Ala Gly Asn Tyr His Leu Val Asp Glu Asn Phe Asp Pro Leu		
440	445	450
Pro Asp Tyr Trp Leu Ser Leu Leu Phe Lys Lys Leu Val Gly Thr		
455	460	465
Lys Val Leu Met Ala Ser Val Gln Gly Ser Lys Arg Arg Lys Leu		
470	475	480
Arg Val Tyr Leu His Cys Thr Asn Thr Asp Asn Pro Arg Tyr Lys		
485	490	495
Glu Gly Asp Leu Thr Leu Tyr Ala Ile Asn Leu His Asn Val Thr		
500	505	510
Lys Tyr Leu Arg Leu Pro Tyr Pro Phe Ser Asn Lys Gln Val Asp		
515	520	525
Lys Tyr Leu Leu Arg Pro Leu Gly Pro His Gly Leu Leu Ser Lys		
530	535	540
Ser Val Gln Leu Asn Gly Leu Thr Leu Lys Met Val Asp Asp Gln		
545	550	555
Thr Leu Pro Pro Leu Met Glu Lys Pro Leu Arg Pro Gly Ser Ser		
560	565	570
Leu Gly Leu Pro Ala Phe Ser Tyr Ser Phe Phe Val Ile Arg Asn		
575	580	585
Ala Lys Val Ala Ala Cys Ile		
590	592	

(2) INFORMATION FOR SEQ ID NO:15:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 1899
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: double

11

(D) TOPOLOGY: linear
(xi) SEQUENCE DESCRIPTION: SEQ ID NO:15

AAA GCG AGC AAG GAA GTA GGA GAG AGC CGG GCA GGC GGG GCG GGG	3
TTG GAT TGG GAG CAG TGG GAG GGA TGC AGA AGA GGA GTG GGA GGG	48
ATG GAG GGC GCA GTG GGA GGG GTG AGG AGG CGT AAC GGG GCG GAG	93
Met Glu Gly Ala Val Gly Gly Val Arg Arg Arg Asn Gly Ala Glu	138
5 10 15	
GAA AGG AGA AAA GGG CGC TGG GGC TCG GCG GGA GGA AGT GCT AGA	183
Glu Arg Arg Lys Gly Arg Trp Gly Ser Ala Gly Gly Ser Ala Arg	
20 25 30	
GCT CTC GAC TCT CCG CTG CGC GGC AGC TGG CGG GGG GAG CAG CCA	228
Ala Leu Asp Ser Pro Leu Arg Gly Ser Trp Arg Gly Glu Gln Pro	
35 40 45	
GGT GAG CCC AAG ATG CTG CTG CGC TCG AAG CCT GCG CTG CCG CCG	273
Gly Glu Pro Lys Met Leu Leu Arg Ser Lys Pro Ala Leu Pro Pro	
50 55 60	
CCG CTG ATG CTG CTG CTC CTG GGG CCG CTG GGT CCC CTc TCC CCT	318
Pro Leu Met Leu Leu Leu Gly Pro Leu Gly Pro Leu Ser Pro	
65 70 75	
GGC GCC CTG CCC CGA CCT GCG CAA GCA CAG GAC GTc GTG GAC CTG	363
Gly Ala Leu Pro Arg Pro Ala Gln Ala Gln Asp Val Val Asp Leu	
80 85 90	
GAC TTc TTC ACC CAG GAG CCG CTG CAC CTG GTG AGC CCC TCG TTC	408
Asp Phe Phe Thr Gln Glu Pro Leu His Leu Val Ser Pro Ser Phe	
95 100 105	
CTG TCC GTC ACC ATT GAC GCC AAC CTG GCC ACG GAC CCG CGG TTC	453
Leu Ser Val Thr Ile Asp Ala Asn Leu Ala Thr Asp Pro Arg Phe	
110 115 120	
CTC ATC CTC CTG GGT TCT CCA AAG CTT CGT ACC TTG GCC AGA GGC	498
Leu Ile Leu Leu Gly Ser Pro Lys Leu Arg Thr Leu Ala Arg Gly	
125 130 135	
TTG TCT CCT GCG TAC CTG AGG TTT GGT GGC ACC AAG ACA GAC TTC	543
Leu Ser Pro Ala Tyr Leu Arg Phe Gly Gly Thr Lys Thr Asp Phe	
140 145 150	
CTA ATT TTC GAT CCC AAG AAG GAA TCA ACC TTT GAA GAG AGA AGT	588
Leu Ile Phe Asp Pro Lys Lys Glu Ser Thr Phe Glu Glu Arg Ser	
155 160 165	
TAC TGG CAA TCT CAA GTC AAC CAG GAT ATT TGC AAA TAT GGA TCC	633
Tyr Trp Gln Ser Gln Val Asn Gln Asp Ile Cys Lys Tyr Gly Ser	
170 175 180	
ATC CCT CCT GAT GTG GAG AAG TTA CGG TTG GAA TGG CCC TAC	678
Ile Pro Pro Asp Val Glu Glu Lys Leu Arg Leu Glu Trp Pro Tyr	
185 190 195	

CAG GAG CAA TTG CTA CTC CGA GAA CAC TAC CAG AAA AAG TTC AAG	723	
Gln Glu Gln Leu Leu Leu Arg Glu His Tyr Gln Lys Lys Phe Lys		
200	205	210
AAC AGC ACC TAC TCA AGA AGC TCT GTA GAT GTG CTA TAC ACT TTT	768	
Asn Ser Thr Tyr Ser Arg Ser Ser Val Asp Val Leu Tyr Thr Phe		
215	220	225
GCA AAC TGC TCA GGA CTG GAC TTG ATC TTT GGC CTA AAT GCG TTA	813	
Ala Asn Cys Ser Gly Leu Asp Leu Ile Phe Gly Leu Asn Ala Leu		
230	235	240
TTA AGA ACA GCA GAT TTG CAG TGG AAC AGT TCT AAT GCT CAG TTG	858	
Leu Arg Thr Ala Asp Leu Gln Trp Asn Ser Ser Asn Ala Gln Leu		
245	250	255
CTC CTG GAC TAC TGC TCT TCC AAG GGG TAT AAC ATT TCT TGG GAA	903	
Leu Leu Asp Tyr Cys Ser Ser Lys Gly Tyr Asn Ile Ser Trp Glu		
260	265	270
CTA GGC AAT GAA CCT AAC AGT TTC CTT AAG AAG GCT GAT ATT TTC	948	
Leu Gly Asn Glu Pro Asn Ser Phe Leu Lys Lys Ala Asp Ile Phe		
275	280	285
ATC AAT GGG TCG CAG TTA GGA GAA GAT TAT ATT CAA TTG CAT AAA	993	
Ile Asn Gly Ser Gln Leu Gly Glu Asp Tyr Ile Gln Leu His Lys		
290	295	300
CTT CTA AGA AAG TCC ACC TTC AAA AAT GCA AAA CTC TAT GGT CCT	1038	
Leu Leu Arg Lys Ser Thr Phe Lys Asn Ala Lys Leu Tyr Gly Pro		
305	310	315
GAT GTT GGT CAG CCT CGA AGA AAG ACG GCT AAG ATG CTG AAG AGC	1083	
Asp Val Gly Gln Pro Arg Arg Lys Thr Ala Lys Met Leu Lys Ser		
320	325	330
TTC CTG AAG GCT GGT GGA GAA GTG ATT GAT TCA GTT ACA TGG CAT	1128	
Phe Leu Lys Ala Gly Gly Glu Val Ile Asp Ser Val Thr Trp His		
335	340	345
CAC TAC TAT TTG AAT GGA CGG ACT GCT ACC AGG GAA GAT TTT CTA	1173	
His Tyr Tyr Leu Asn Gly Arg Thr Ala Thr Arg Glu Asp Phe Leu		
350	355	360
AAC CCT GAT GTC TTG GAC ATT TTT ATT TCA TCT GTG CAA AAA GTT	1218	
Asn Pro Asp Val Leu Asp Ile Phe Ile Ser Ser Val Gln Lys Val		
365	370	375
TTC CAG GTG GTT GAG AGC ACC AGG CCT GGC AAG AAG GTC TGG TTA	1263	
Phe Gln Val Val Glu Ser Thr Arg Pro Gly Lys Lys Val Trp Leu		
380	385	390
GGA GAA ACA AGC TCT GCA TAT GGA GGC GGA GCG CCC TTG CTA TCC	1308	
Gly Glu Thr Ser Ser Ala Tyr Gly Gly Ala Pro Leu Leu Ser		
395	400	405

13

GAC ACC TTT GCA GCT GGC TTT ATG TGG CTG GAT AAA TTG GGC CTG	1353	
Asp Thr Phe Ala Ala Gly Phe Met Trp Leu Asp Lys Leu Gly Leu		
410	415	420
TCA GCC CGA ATG GGA ATA gAA GTG GTG ATG AGG CAA GTA TTC TTT	1398	
Ser Ala Arg Met Gly Ile Glu Val Val Met Arg Gln Val Phe Phe		
425	430	435
GGA GCA GGA AAC TAC CAT TTA GTG GAT GAA AAC TTC GAT CCT TTA	1443	
Gly Ala Gly Asn Tyr His Leu Val Asp Glu Asn Phe Asp Pro Leu		
440	445	450
CCT GAT TAT TGG CTA TCT CTT CTG TTC AAG AAA TTG GTG GGC ACC	1488	
Pro Asp Tyr Trp Leu Ser Leu Leu Phe Lys Lys Leu Val Gly Thr		
455	460	465
AAG GTG TTA ATG GCA AGC GTG CAA GGT TCA AAG AGA AGG AAG CTT	1533	
Lys Val Leu Met Ala Ser Val Gln Gly Ser Lys Arg Arg Lys Leu		
470	475	480
CGA GTA TAC CTT CAT TGC ACA AAC ACT GAC AAT CCA AGG TAT AAA	1578	
Arg Val Tyr Leu His Cys Thr Asn Thr Asp Asn Pro Arg Tyr Lys		
485	490	495
GAA GGA GAT TTA ACT CTG TAT GCC ATA AAC CTC CAT AAC GTC ACC	1623	
Glu Gly Asp Leu Thr Leu Tyr Ala Ile Asn Leu His Asn Val Thr		
500	505	510
AAG TAC TTG CGG TTA CCC TAT CCT TTT TCT AAC AAG CAA GTG GAT	1668	
Lys Tyr Leu Arg Leu Pro Tyr Pro Phe Ser Asn Lys Gln Val Asp		
515	520	525
AAA TAC CTT CTA AGA CCT TTG GGA CCT CAT GGA TTA CTT TCC AAA	1713	
Lys Tyr Leu Leu Arg Pro Leu Gly Pro His Gly Leu Leu Ser Lys		
530	535	540
TCT GTC CAA CTC AAT GGT CTA ACT CTA AAG ATG GTG GAT GAT CAA	1758	
Ser Val Gln Leu Asn Gly Leu Thr Leu Lys Met Val Asp Asp Gln		
545	550	555
ACC TTG CCA CCT TTA ATG GAA AAA CCT CTC CGG CCA GGA AGT TCA	1803	
Thr Leu Pro Pro Leu Met Glu Lys Pro Leu Arg Pro Gly Ser Ser		
560	565	570
CTG GGC TTG CCA GCT TTC TCA TAT AGT TTT TTT GTG ATA AGA AAT	1848	
Leu Gly Leu Pro Ala Phe Ser Tyr Ser Phe Phe Val Ile Arg Asn		
575	580	585
GCC AAA GTT GCT GCT TGC ATC TGA AAA TAA AAT ATA CTA GTC CTG	1893	
Ala Lys Val Ala Ala Cys Ile		
590	592	
ACA CTG	1899	

(2) INFORMATION FOR SEQ ID NO:16:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 594

(B) TYPE: nucleic acid

14

(C) STRANDEDNESS: double
 (D) TOPOLOGY: linear
 (xi) SEQUENCE DESCRIPTION: SEQ ID NO:16

ATTACTATAG GGCACGCGTG GTCGACGGCC CGGGCTGGTA TTGTCTTAAT GAGAAGTTGA 60
 TAAAGAATTT TGGGTGGTTG ATCTCTTCC AGCTGCAGTT TAGCGTATGC TGAGGCCAGA 120
 TTTTTCAAGG CAAAAGTAAA ATACCTGAGA AACTGCCTGG CCAGAGGACA ATCAGATTTT 180
 GGCTGGCTCA AGTGACAAGC AAGTGTGTTT AAGCTAGATG GGAGAGGAAG GGATGAATAC 240
 TCCATTGGAG GCTTTACTCG AGGGTCAGAG GGATACCCGG CGCCATCAGA ATGGGATCTG 300
 GGAGTCGGAA ACGCTGGTT CCCACGAGAG CGCGCAGAAC ACGTGCGTCA GGAAGCCTGG 360
 TCCGGGATGC CCAGCGCTGC TCCCCGGCG CTCCCTCCCG GGCCTCCCTC CCCAGGCCTC 420
 CGGGGCGCTT GGATCCCGCA CATCTCCGCA CCCTTCAAGT GGGTGTGGGT GATTTCGTAA 480
 GTGAACGTGA CGGCCACCGG GGGGAAAGCG AGCAAGGAAG TAGGAGAGAG CGGGGCAGGC 540
 GGGGGGGGGT TGGATTGGGA GCAGTGGGAG GGATGCAGAA GAGGAGTGAGG AGGG 594

(2) INFORMATION FOR SEQ ID NO:17:

(i) SEQUENCE CHARACTERISTICS:
 (A) LENGTH: 21
 (B) TYPE: nucleic acid
 (C) STRANDEDNESS: single
 (D) TOPOLOGY: linear
 (xi) SEQUENCE DESCRIPTION: SEQ ID NO:17
 CCCCAGGAGC AGCAGCATCA G 21

(2) INFORMATION FOR SEQ ID NO:18:

(i) SEQUENCE CHARACTERISTICS:
 (A) LENGTH: 21
 (B) TYPE: nucleic acid
 (C) STRANDEDNESS: single
 (D) TOPOLOGY: linear
 (xi) SEQUENCE DESCRIPTION: SEQ ID NO:18
 AGGCTTCGAG CGCAGCAGCA T 21

(2) INFORMATION FOR SEQ ID NO:19:

(i) SEQUENCE CHARACTERISTICS:
 (A) LENGTH: 22
 (B) TYPE: nucleic acid
 (C) STRANDEDNESS: single
 (D) TOPOLOGY: linear
 (xi) SEQUENCE DESCRIPTION: SEQ ID NO:19
 GTAATACGAC TCACTATAGG GC 22

(2) INFORMATION FOR SEQ ID NO:20:

(i) SEQUENCE CHARACTERISTICS:
 (A) LENGTH: 19
 (B) TYPE: nucleic acid
 (C) STRANDEDNESS: single
 (D) TOPOLOGY: linear
 (xi) SEQUENCE DESCRIPTION: SEQ ID NO:20
 ACTATAGGGC ACGCGTGGT 19

(2) INFORMATION FOR SEQ ID NO:21:

(i) SEQUENCE CHARACTERISTICS:
 (A) LENGTH: 21
 (B) TYPE: nucleic acid
 (C) STRANDEDNESS: single
 (D) TOPOLOGY: linear

15

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:21
CTTGGGCTCA CCTGGCTGCT C 21

(2) INFORMATION FOR SEQ ID NO:22:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 23
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:22
AGCTCTGTAG ATGTGCTATA CAC 23

(2) INFORMATION FOR SEQ ID NO:23:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 22
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:23
GCATCTTAGC CGTCTTTCTT CG 22

(2) INFORMATION FOR SEQ ID NO:24:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 23
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:24
GAGCAGCCAG GTGAGCCCAA GAT 23

(2) INFORMATION FOR SEQ ID NO:25:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 23
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:25
TTCGATCCCA AGAAGGAATC AAC 23

(2) INFORMATION FOR SEQ ID NO:26:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 23
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:26
AGCTCTGTAG ATGTGCTATA CAC 23

(2) INFORMATION FOR SEQ ID NO:27:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 24
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:27
TCAGATGCAA GCAGCAACTT TGGC 24

(2) INFORMATION FOR SEQ ID NO:28:
(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 22
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear
(xi) SEQUENCE DESCRIPTION: SEQ ID NO:28
GCATCTTAGC CGTCTTCTT CG 22

(2) INFORMATION FOR SEQ ID NO:29:
(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 24
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear
(xi) SEQUENCE DESCRIPTION: SEQ ID NO:29
GTAGTGATGC CATGTAAC TG AATC 24

(2) INFORMATION FOR SEQ ID NO:30:
(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 22
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear
(xi) SEQUENCE DESCRIPTION: SEQ ID NO:30
AGGCACCCCTA GAGATGTTCC AG 22

(2) INFORMATION FOR SEQ ID NO:31:
(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 24
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear
(xi) SEQUENCE DESCRIPTION: SEQ ID NO:31
GAAGATTTCT GTTTCCATGA CGTG 24

(2) INFORMATION FOR SEQ ID NO:32:
(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 25
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear
(xi) SEQUENCE DESCRIPTION: SEQ ID NO:32
CCACACTGAA TGTAATACTG AAGTG 25

(2) INFORMATION FOR SEQ ID NO:33:
(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 22
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear
(xi) SEQUENCE DESCRIPTION: SEQ ID NO:33
CGAAGCTCTG GAACTCGGCA AG 22

(2) INFORMATION FOR SEQ ID NO:34:
(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 22

17

(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:34

GCCAGCTGCA AAGGTGTTGG AC 22

(2) INFORMATION FOR SEQ ID NO:35:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 23
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:35

AACACCTGCC TCATCACGAC TTC 23

(2) INFORMATION FOR SEQ ID NO:36:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 22
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:36

GCCAGGGCTGG CGTCGATGGT GA 22

(2) INFORMATION FOR SEQ ID NO:37:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 22
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:37

GTCGATGGTG ATGGACAGGA AC 22

(2) INFORMATION FOR SEQ ID NO:38:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 22
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:38

GTAATACGAC TCACTATAGG GC 22

(2) INFORMATION FOR SEQ ID NO:39:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 19
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:39

ACTATAGGGC ACGCGTGGT 19

(2) INFORMATION FOR SEQ ID NO:40:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 27
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:40
CCATCCTAAT ACGACTCACT ATAGGGC 27

(2) INFORMATION FOR SEQ ID NO:41:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 23
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:41

ACTCACTATA GGGCTCGAGC GGC 23

(2) INFORMATION FOR SEQ ID NO:42:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 44848
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: double
- (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:42

GGATCTTGGC	TCACTGCAAT	CTCTGCCCTC	CATGCAATT	TTATGCATCA	50
GCCTCCTGAG	TAGCTTGAT	TATAAGGTCTG	CGCCACCACT	CCTGGCTACA	100
CCATGTTGCC	CAGGCTGGTC	TTGAACTCTT	GGGCTCTAGT	GATCCACCCG	150
CCCTGGCCTC	CCAAAGTGC	GGGATTACAG	GTGTGAGCCA	TCACACCCGG	200
CCCCCGCTTT	CCATATTAGT	AACTCAGATC	TAGACCAACAA	GGATGCACTA	250
TTTAGAAAC	TTGCAATGGT	CCACTTTCA	AATCACCCAA	ACATGTTAAA	300
GAAATTGGTA	TGACTGGCA	TGGCACAGTG	GCTCATGCCT	GCAATCCTAG	350
CATTTTGTA	GGCTGAGACG	GGCAGATCAC	GAGGTGAGGA	GATTGAGACC	400
ATCCTGACAG	ACATGGTAA	ATCCCCATCTC	TACTAAAAAT	ACAAAACAAAT	450
TAGCCGGGGG	TGATGGCAGG	CCCCCTGTAGT	CCCAAGCTACT	CGGGAGGCTG	500
AGGCAGGAGA	ATGGCGTGA	TCCAGGAGGC	AGAGCTTGCA	GTGAGCCGAG	550
ATGGTGCCAC	TGCACTCCAG	CCTGGGCAC	AGAGCGAGAC	TCCGTCTCAA	600
AAAAAA	AAAGAAAGAA	ATTGGTATGA	CTGTTGACTC	ACAACAGGAG	650
TCAGGGGCAT	GGGGTGGGGT	GTAAGATTA	TGTCATGACA	AATGTGAAA	700
AGAAAACCTCT	GTTCCTTCAA	CTCCACGTCT	GCTACCATAT	TATTACACTC	750
TTCTGGTAGT	GTGGTGTGTTA	TGTGTGAAATT	TTTTTCATA	TGTATACAGT	800
AATTGTAGGA	TATGAACTCTG	ATTCCTAGTTG	CAAAACTCAC	TATGAGCTTA	850
GCTTTAAAGT	TGCTTAAGAA	TAGGTAGATC	TATGCAAAATA	ATGATAATT	900
TTATTATTAT	TTTAAGAGAG	GGTCTCACTT	TGTCACCCAG	GCTGGAGTGC	950
AGTGGTGTGA	TTAAGGGTCA	CTGCAACCTC	CACCTCCCAG	GCTCAAATAA	1000
ACCTCCCACC	TCAGCCTCCC	CACTAGCTGG	AACCACAGGC	ACGGGCCACC	1050
ACGCCTGGCT	AATTTTTTGT	ATTTTTTGTG	GAGATGGGGT	TTCATCATGT	1100
TGCCCCAGGT	GTTCTTGAAT	TCCCTGGCTC	AAGCAATCT	CCACCTTGG	1150
CCTCCCCAAA	TGCTGGCATC	ACAGGCATGA	TGGCATCACT	GGCATCACAT	1200
ACCATGCTG	GCCTGATTTA	TGCAAATTAG	ATATGCATTT	AAAATAATC	1250
TATTTTTATT	TGTTGCTTCA	TTGGTGTAC	AATCTCAAGT	GGAAAAATCT	1300
AAGGGTTTTG	GTGTTATTG	CTTACTCAAC	CAATATTAT	TAGACTCTTA	1350
CTAAGCACCA	ACATGATCAC	ATGCCTGAGC	TATGGCTAGC	ATAGCGTGTG	1400
AGACAAACCT	AATCTCTGTT	TTGGTGGAGC	ATATAATCTA	GTAGATGAAG	1450
CCAATATGTA	GCAACATCAC	AATACTAAC	AATTGAGGAT	GCTACGAGAG	1500
TGCTAACAA	ATTGAGGATG	CTACGAGAGT	GTCTAACAAA	TTGAGGATGC	1550
TATGAGAGTG	TGTCATGGAG	AGCTGCTTG	AGATTGAGAG	AAAGCTTCT	1600
TGAGGGAAAGT	TACATTTCA	CTGAAACACA	CTGCCATCTG	CTCGAGGTT	1650
TGTAATCTGCA	TTCACATCCC	GATTCTGACA	CTTCACATCC	CGATTCTGAC	1700
ACTTCACCCA	GTTACTGTCT	CAGAGCTTGG	GTCCCGCATGT	GTAAAACAG	1750
GACAGTATGC	ACTTGGCAGG	CTTGTGAGAA	GGGAAGAGAA	CACAAGTAAA	1800
GCACCTGTAT	CAGGCATACA	GTAGGCACTA	AGCGTGCAT	GCTTGCTATG	1850
ATTATACATC	AGTGTAAAGCA	TCAAGGAAAA	GCTGAAGAAA	AGTCTGACCA	1900
ACAGCGAAAG	ATAAATGCGC	AGAGGAGAAA	TTTGGCAAAAG	GCTCCAAATT	1950
CAGGGGCACT	CCGTACTCTA	CACTTTGTAT	GGGGGCTTC	GGTCCTGAGT	2000
TCCAGACATT	GGAGCAACTA	ACCCTTAAC	ATTGCTAAAT	ATTGTCTTAA	2050
TGAGAAAGTTG	ATAAGAATT	TTGGGTGGTT	GATCTCTTT	CAGCTGCACT	2100
TTAGCGTATG	CTGAGGCCAG	ATTTTTCA	GCAAAAGTAA	AATACCTGAG	2150
AAACTGCTG	GCCAGAGAC	AATCAGATTT	TGGCTGGCTC	AAAGTGACAAG	2200
CAAGTCTTA	TAAGCTAGAT	GGGAGAGGAA	GGGATGAATA	CTCCATTGGA	2250
GGTTTTACTC	GAGGGTCAGA	GGGATACCCG	GCGCCATCAG	AATGGGATCT	2300
GGGAGTCGGA	AACGCTGGGT	TCCCACGAGA	GCGCGCAGAA	CACGTGGCTC	2350
AGGAAGCCTG	GTCCGGGATG	CCCAGCGCTG	CTCCCCGGGC	GCTCCCTCCCC	2400

GGGCGCTCCT	CCCCAGGCCT	CCGGGGCGCT	TGGATCCCGG	CCATCTCCGC	2450
ACCCCTTCAG	TGGGTGTGGG	TGATTCGTA	AGTGAACGTG	ACCGCCACCG	2500
AGGGGAAAGC	GAGCAAGGAA	CTAGGAGAGA	GCCGGCGGAG	CGGGGGGGGG	2550
TTGGATTGGG	AGCACTGGGA	GGGATCAGA	AGAGGAGTGG	GAGGGATGGA	2600
GGGCCAGTG	GGAGGGGTGA	GGAGGCGTAA	CGGGGGCGAG	GAAAGGGAGAA	2650
AAGGGCGCTG	GGGCTCGGGC	GGAGGAAGTG	CTAGAGCTCT	CGACTCTCCG	2700
CTGCCGGCA	GCTGGCGGGG	GGAGCAGCCA	GGTGAGGCCA	AGATGCTGCT	2750
GCGCTCGAAG	CCTGCGCTGC	GGCGCCGCT	GATGCTGCTG	CTCCCTGGGC	2800
CGCTGGGTCC	CCTCTCCCT	GGGCCCTGTC	CCCGACCTGC	GCAAGCACAG	2850
GACGTCGTGG	ACCTGGACTT	CTTCACCCAG	GAGCCGCTGC	ACCTGGTGAG	2900
CCCTCTGTT	CTGTCGTC	CCATTGACGC	CAACCTGGCC	ACGGACCCGC	2950
GGTTCCTCAT	CCTCTGGGG	TAAGGCCAG	CCTCTGGTC	CTGTCCTCTT	3000
TCCCTGTCCTC	CTGACACCTA	TGTCTGGCCC	GCCAGCGGTG	CTCCCTCTTT	3050
TGCGCGGAAA	CAACTTCACA	CCGGAACCTC	CCCGCCCTGTC	TCTCCCCACC	3100
CCACTTCCCG	CCTCTCATTC	TCCCTCTCCC	TCCCTTAAC	TCAGACCCCA	3150
AACCGCTTTT	TGGGGGTAT	CATTTAAAAA	ATAGATTTAG	GGTTTACAAG	3200
TGCAGTTCTG	TTCCATGGG	ATATTGATT	GTGGTGGCAT	CTGGGCTCTT	3250
AGTGTAACTG	TCACCCGAAT	GTTGTACATT	GTATCTAATA	GGTATTCTCT	3300
CATCCCTCAT	CCCTCTCCC	CCCTCCCAAC	TTTTGGAGTC	TCCAGTGTCT	3350
ACTATTCCAC	TAAGTCCATG	TGTACACATT	GTGAGGCC	CACTCTAAAT	3400
GAGCCTTTT	GTTTCAATTCA	TTCTGTAAGT	GTTGAATAGG	CACCCACCTAA	3450
GTCAGGTAT	AAGTGGAAAT	TTGAAAAGA	AACTGCCAC	TTGCCCCAGT	3500
ACTTCCCTAG	CCAAGAGGAG	GGAAACCAAG	CAGGTGCACC	TGAAGGCCCTG	3550
TGAGTGTCTG	ATTGCTGTG	CAAGTGTAGGA	CAAGTAAGAT	TGTGCATAGC	3600
CTTCTGTATT	TAAGACTCTG	TTAGGAAGAT	TTCTCTTTCT	TTCTCTTTCT	3650
TTTTCTTTT	TCTTTCTT	TTTTTTTTA	GGCAGATGAA	AAGGGCGTCA	3700
CAGAACAGGA	ATAAAAATCT	AAATATTCAA	TAATGAGAC	CTAGGAGACT	3750
ACTGCAGTGA	CTTACAAAGT	CCTAATAAAAA	AGATGCTCT	CCAAAATGGG	3800
GCTGAAAT	GTGGTGTGTC	CTTATCAGCT	CTAAGTTTTT	TCCCTACCTG	3850
AGAAAAGAGG	AACCTGATG	AGGTTCAAGG	CTCCGTCCCC	ATGAATGAG	3900
GCTGACTCCA	AGATGGGGAG	CTACAGGGAC	AATCCCAAGG	CTTCTAGGCC	3950
TCTTATTCTG	GCCCTGGGG	CCTCCAGAGA	TGGCCACATC	TTGACCAGCC	4000
CAGATAGAGG	GAAAGATCAC	CATTATCTCA	CCTCTGTGTC	AAATACCTAG	4050
ATGCTGTCT	CCCTGAGCCC	ACACTATAGT	TGCCAGCGTC	AATTAAATGG	4100
GTAGTGTACT	GGTTAAGAGA	TGGACAGACC	ATCTGGCTT	GACTCTCAGC	4150
TCTGGCAAAG	ATGAGTGA	TGGTTTTCC	ATATCTCTT	GCACACACAA	4200
CCTTGATTT	TTCAGCTGTA	GAATGGAATT	TCTCAAGCTT	GCCTCAAGGA	4250
TTATTGCCCC	AGGATTGAT	GATATGGTAA	GAGCTTCTCA	GTGTTTGACC	4300
CATAGTAAGT	GTGGACGTT	TCAAACGAAT	TGTTTCTTT	TAGGACATGG	4350
TGAGCATTG	GTAGCCATTTC	ACCGGTTTC	TGTTTCTTTG	GATCATAGTT	4400
AACCTCTCCT	TTTCTCTCTG	GCACCTACAAT	TTTCTGGTGG	GGAAAGATCC	4450
TTACTTTCTG	CCCTCTCCCT	TAAGGATAGG	AACTGATAC	TAGGCAGCAA	4500
CTAGTGGGG	GATAGGAATT	TCTCAAGCTT	GCCTCAAGGA	4550	
TCCAGATCAC	AGGACCCAG	TCTTAGCTT	CTGGGGTGTG	GGGTGGGGGG	4600
GGCGGGTTAC	TGAACATGGG	TATGAAGTAG	ATGTCATTT	ACTGAAATGT	4650
GAGGACCTGA	GGCCTCTCT	ATTGCTGTAG	CCAGCATATT	CCCCAACCTC	4700
TCCCCAAGA	AGGACAGATG	GGGGTTCCCC	CCTGGAGTAA	CAGGTCCAAA	4750
AGAAAAAACAA	TACAGTGGGA	CTTCCAGGAT	CTGGGGCTGA	TCAACCCAGA	4800
GTCAGCTCC	CCGCAATTGA	CTAACACCCCC	CCTAACACGT	AGAAAATCCA	4850
ATCTGCAATT	TAGTGGAGAT	GATACTTTA	TTCTCTTAA	ATACATCTCT	4900
TCATTCTCCA	GAGCCACCTT	TTTCTCCCTC	CTCTGCACCT	TTTTGTTAAA	4950
GACTGGAGTA	TAATGAAATA	CCAAGAGAGC	ATAACATGTG	ATACATAAAA	5000
CTTTTTCTC	GGTTTACAAA	ACAGTTCTT	CTTGTCTCAT	CGTGTCTCTC	5050
TCCAAGGCTG	GCTGCTGTCT	GTTCCAGCCC	GCTTCGCTTG	GAGAGGCCAT	5100
CTGCCATACC	TGCTCCCCAG	ACGCATCGAC	AAGCACACCC	AGAGTGTAT	5150
CTGCTAAGAC	CTAAAAGAG	GAGGAACCCC	CTCTCTCAT	CTAACACCTA	5200
GCTTCTAAAT	TAGAGTGTGA	GGGTCCCAT	CCCCAGGAGG	GGCACAGGGC	5250
CCAAACAGCC	CAGCCATCTC	AGAACAGAAC	ACTAACGTTT	GTAGGGTCC	5300
ACAGTAGAGG	AGAGTAAGAC	GGCTGTTGTT	TAATTCTTA	CAGTTCTCA	5350
AAAGTGAAGA	TGTGTGGGG	GGATGGCAAG	AGCTGAGCAC	ACGAAAGCTG	5400
AGGAATAAG	AGGACACAAA	CAGCTGACAC	TTCTCTAGTT	5450	
CTTGTCTATT	GCCTGGCCCT	GTTCTAAGCA	CCTCTCTAGGT	ATTAATCCAT	5500
TTAGTCTTGG	CTACAACACT	GTGACTAAT	AGTTTTGTCA	CCCCATTTT	5550
AAAAATGAAG	AAAGTGAAGG	TCAGGGAGGT	TAAGTAACCT	GGCCACAGTT	5600
TGAAACTAGA	CTCTGATCAC	ATGAGATAAT	AGTGCCTATA	AAAAGGGAAA	5650
GCAGATTATA	TTTTTTAAAG	GAAGAGAGT	AGGATATGGT	AGAAAAAGAT	5700
TGTTTGAAA	GGAATTGAGA	GATTGATATA	ATGAAAAGAA	GCATTCACAT	5750
GAGAGTAACA	GTATCAGGGC	CCAAACCTTC	ATCTAAGGA	CTTCAAGAGAG	5800
GCTTAAGCAA	ACTTAGTCAC	TGGCGTGGTT	CTAGTCCTCA	TGATGGCAAA	5850
TACATTGTGT	ACAGCCCCAC	TCCACACAAA	ACTTAAATAC	CAATGATAGA	5900
GCAATCTAAA	ATTGAAAGA	AAAAATCTT	CAATTTGTG	TCTTCCCAGA	5950
GGGACTTAAAT	CAAGAAACCA	ATCAAAATAC	TTCTCTAAGCC	TAACTGTGTG	6000
CAGAACTCCA	AAGAGAGGCC	AGCCCTAAAT	CAACACTGTC	CAATGGAAAT	6050
ATAATATAAT	GTGGGCTCA	TATGCAAGGT	CATATGTAAT	TTTAAATTCTT	6100
CTAGTAGCCA	TATTTAAAGAAG	ACAACTGAAA	TTAATTCTTAA		6150

TAATTTTATT	TAGTTCAATA	GATCCAAAAT	GTTCCTCAG	CATGTAATCA	6200
ATATAAAAAT	ATTAATGAGG	TATTTATTAT	TCCTTTCTC	AAACCAAGTC	6250
TATTCTATAA	TCTGGCGTGT	ATTATTACCA	GCACCTCTCA	GACTATATT	6300
CTTCTTTCT	TTTTTTTC	CGAGACAAATT	TTGCTCTTGT	CACCCAAAGCT	6350
AGAGTACAAT	GGCGTTACCT	CGGCTCACTG	CAACCTCCCG	CTCCCCGGTT	6400
CAAGTTATT	TCCTGCCTCA	GTCTCCCAAG	TAGCTGGGAC	TAGAGGCATG	6450
CACCAACCACG	CCTGGCTAAT	TGTGTATT	TAGTAGAGAC	AGGGTTTCAC	6500
CATGTTGGCC	AGGCTAATCT	CAAACCTCTG	AGCTCAGGTG	ATATGCCAC	6550
CTCGGCTCTC	CAAAAGTGTG	GGTACACAGG	CGTGAGGCC	TGCAACCGGC	6600
CTCAGATTAA	CTATATTTC	AGCGTTCACT	AGCCACATGT	AGCTAGTGT	6650
ATGCTAGTGG	ACAGTACAGA	TCTGCATTTC	AATTAAGACA	CGTATACAAG	6700
CATAGTTCA	TAATGCAACG	TAAAAAAAG	TATAGTGTG	AGTCGGTGGT	6750
AGAAATCTCA	AATACTGCA	AGCAAAAGTG	GTACAAACAG	CAATCTCAGT	6800
GATAATGCAA	CCATGCTTC	TTTCATTG	AATTGCTTA	TTTTCTTCA	6850
GCAAAGTTCA	TCCATTTTG	CCAATTCAAT	AAATATTAC	TGATAAAAAC	6900
TTCAATATT	AGATTCCTGC	ATCTTCATAG	ACAGAGTTGC	TTTTCACATT	6950
TAGAAAATTA	CTTATCAATG	TTAAACACAC	GTTTGATAAA	CCAGTGTGG	7000
AAAAGGGTGC	AGACTCCCCA	TGTGCTTATT	GATGGCAGAA	ATATTCACAG	7050
CCAAAGGGAA	ACAAAGGGCT	GGGGACAATC	ACACACCTCA	TGTCTCTAA	7100
CTCCTGGGAA	GTGCTGTC	TCTGATTGAG	CTCTTATTAT	TGCCCTCCCC	7150
ACTAACCTCG	TCCACTGTG	CCTGGAGCCC	TTTGAGGTTG	TACCTGCTCT	7200
GTCTCCCTCA	CGAAATATCT	CTCTACCTC	CTTGCTTCAAG	CTACAACCTG	7250
GCTATTCTCT	GATGACACTG	TCTTCCCTGT	AGCCCTTTG	AGTAATGGCT	7300
GCATATTCTC	CCATAGTCCA	GTTCTTTCC	TGTTCTCCAG	TCTGGCTTCT	7350
GGATGACAGC	CCACTAGTT	GAACCTCATA	CTGCTATAGT	TCAAGTCCCCT	7400
TTGACTTGT	TACCTTGGG	AAATTACCTC	CTTTGTTCA	GGTTCTTGT	7450
TTGAAAATG	ACGATAATAA	TGCCATTG	TTCACTGGGT	TATTTTGAAA	7500
TTGAGTGA	GAAGGCGGGT	AGCTTCCCTA	CACGCTCAGT	GTAGACTAGC	7550
CTGATGTCGA	TTACGGGTGA	TGCCATGACT	CAGTGTGTT	TCCTCATCTC	7600
CACATCTGGC	TCTCATCCAG	TGCTCTGCT	TACGGCACTC	TGTCCTCCCTC	7650
TTACTTACTC	CCCCCTTATT	ACTGAAGAT	GGCACTGATC	TCACAGTTTC	7700
CTCTCCACTT	CCTAGTCTCA	CCATCATCCT	AGATGACTTC	AAGTCACCTA	7750
GATAAACTGT	CTCAGTTCT	TCACTCACAT	TTTTTATAA	CAGATAATGT	7800
TACACTCAAG	TTGTAACAGA	ACCAGCTTAT	CCAGCTCATG	AAATGTATGC	7850
ATTTCATCTC	AACTCTGTAT	TGAGTCAT	CTCTGGGTA	TCTGGAAATC	7900
AGCCATGGT	AGAATATTAA	CCATGGAAT	TGCAAAATAC	AAAAAGCAG	7950
AGCACCTTTT	TTTCTGAGAG	CCAGACCATA	GCTCTCTAC	TCCATAGCAC	8000
CCATCATAAC	AAATTCTTAA	TACCTCCACT	GAACAGCTTC	TTCTCTCTC	8050
TACTCTCTC	ATATCTGATT	TGAGCTTCTT	AATTATCAT	GTGAACCACT	8100
CTTGTAAATAA	TAACCCAAA	TCCCTGTCTC	ATTGTTCTTC	CTGCTAAAT	8150
ACTAAACCTG	GTTCAGTCCA	ACCATATT	CTCTCTTGG	AATCTACAGG	8200
GTGGCCAAA	AAACCTGGAA	TGGAAAATA	TTACTTATA	ATTTTAATGT	8250
ATTTAAATAA	GCATTTAA	TGCTTCATT	CCAGTCTCAG	TGGCCACCCCT	8300
GTATAGCTGG	GCTATTGAGC	TCTTGCAGGA	GGAGGGAGTG	GACAGTCCTC	8350
CAGCCACACA	GAETGATGTT	GCACCAAACA	TTTTTAGCT	TCCAGACTTC	8400
CCCTGGCCCTT	AGTGTGACCTC	TTAACCTCTC	ATTTCCTCTG	CTTCACATT	8450
CTCTACTTTT	TTAAATCTC	TGACTCCACC	TTCACCTTAT	CATTCTTAGC	8500
ACATGACCAT	ACTTCTGCTT	CCAAAGAAA	ATGAGCAATT	ACTTCCTTT	8550
CCTTTTCCTC	CTGTATCAA	ATCTGCAGAC	ATGTCATGCC	TAAGTCAGC	8600
TTTCTCTCTT	TCTCTGATCT	CACTCTGCTT	CTTCCATTTC	TGCCCTGAAT	8650
CCCGTCCCCCT	CCCCAACCCC	CAAGGACTTC	GCTCTATCAG	TCACCTCTTC	8700
CCTCTCTCTG	ATCTTCAC	CCTCCCATTT	TACTGGCTTC	TCCTCAAGC	8750
CTTCCCCCAA	GCCTTCTCCA	TCTCAATTAC	CTCCTCGCAC	ATGCCCTCTC	8800
AGAAAACCACC	CGTCTTCTC	CCTCCCTCG	GCAGCTGTT	CTTCCGTTC	8850
TGCCCTCATG	ATGCAACCAT	CATTGTCGA	CTAAAATCAA	TCTCTCCGAC	8900
ATCATCAATG	GCCTTCTT	GTTGGAAAC	CTAATAACAA	CTTATCTTA	8950
TTTGGCTTTT	GTTATGGGTT	GAATGAGGTT	ACCCGAAAT	CCATATTAGA	9000
AGTCCTAAC	CCCAGTACCT	CAGAATGTGA	CTTATTGTTG	GAATAGGGTC	9050
ATTGCAAGACG	TTATTAGTTA	GGATGAGGTC	ATACTGGAAT	GTGATGGGCT	9100
GCTTATCTAA	TATGACTGAT	GTCCTTATAA	CAAGGAGAAA	TTTGGAGACA	9150
GACACGCCACA	TAGGGAGAAAT	ACCATGTGAT	GACAGGAGTT	ATGAGTTGG	9200
AGTCAAAAG	CTATGGGAAC	TTAGGAGAAA	GACCTGGAAC	AAATCTTTC	9250
CTGCGCTCTAG	AGAGGGAGTA	TGGCCCTGCC	ACTACCTTGA	ATTCAACGTT	9300
TCGGCTTTTC	AAAATCTGAA	GACAATACAT	TTCTGTTGTT	CAAACCAATT	9350
AGTTGCACTG	ACTCTGCGAC	TGCAGCCCTA	ACAAACTAAT	ACAGTCTCTT	9400
GGAGGCATT	GGCAAGGTTG	ACAATGGAAG	CACTTTCTTA	CCCCTTAGG	9450
TCTGTGCGCT	TTCTTGTG	GGGGTGTGTT	CTAACAAATT	CTCTCCATCT	9500
CTCTCTCTCT	AGTTTGTCTT	AAACATTGCT	GTTCTTCAGA	CTTCTGACCT	9550
AGGCCTTCTT	TTCACTTCAC	ATATTCCCT	GGGTGGTCTC	ACCCACTTCC	9600
AGAAATTACT	AAATTACTG	CTCATGCACT	ACTGTCGTTG	AAACTGTTA	9650
ACAACCTGGCT	CTCTGGGAAG	AGGGGAGACT	GGTTGATGTT	TTTGCTGAT	9700
TTCTGTGGTG	AAATTAATCTC	CTCCATGGCC	AATTCCAAAC	TGCAACAGT	9750
TTAACAACTG	GCTCACAAAT	TTTCTCCAAA	TTAACATTT	GGCTTTCACA	9800
GGCCAAACAC	GTGGTACAGC	CAACTCCAGC	ACACCTCTG	TTTGTGTC	9850
GAGAGAAGTA	ACTTATT	GTACAAAAGG	AAAATAAAA	ACACCTGCAG	9900

GCCCCCTTTT	TTCCTTAAC	AAACTGCTCT	AGAAATAGAA	TAGCTGAAGC	9950
TTCTTTATG	CATTGATCTG	TTATTTCCAT	GTCACTGTGG	TGGTGGGATT	10000
ATTTTCTT	TATTTTCTT	GTATATGGTT	GAAATACTGT	ACCTTTGATC	10050
AGTTTAGT	TTATGGCAT	TTTGCAACCC	ATATTAATAC	TAGTTTTGT	10100
CAGAGGGCGT	CAATATTATT	TTCTCAAAAC	AAGAAAATAT	TTCATTTGCAA	10150
AGGAGACAAA	CAAAAAGGTC	CTTAATACCA	AAACTTTGAA	ATGTGATTC	10200
TTGACTTTGG	CAGTGTCCAA	GTGGTAAAC	CAAACAGTAT	TGGTTTTCTA	10250
TTTGTGTCAG	GAAAGTCTT	GTCTGGCAGC	GACTTACCC	TACATCAGGC	10300
GGGCTTGCT	CATTCACTCA	CTTAAGTATT	TATTAACAC	CAGCGGTGTG	10350
CCAAGTACTT	ATCTAGGTAT	CGGGTAGATT	CTGATAAGTC	AGTCAGGTCC	10400
CTGCTCTCAG	GGAGCTTGCA	GCAGAGATGG	GGGCTGCAAT	AGAGAGTAAG	10450
CCAAGGAAAT	GAAAAGGAA	TGTGATTTCA	GAGAGTGATG	AATGCTATGA	10500
AGAAAATGAA	GGCAGCGCAG	TGTGATGGAG	AGTGACCCAA	GGTGGTACAG	10550
TTTGTACCTC	TAAGGACAC	ACTGTGACCC	AGGTCACTCA	CAGATGCCCG	10600
TCATGTGATG	CCACAGCAAC	TTTTCCAGGT	GCTCGTTCC	TCCCACCTTC	10650
CAGTCTCTTG	CCCAGCCCG	ACTGCTTACA	AATAACAGTC	GAGGAATCTA	10700
AATGAGGTTTC	CTCTATCATC	AAACCCAATC	AAAATGCCAA	GGAACAGAAAT	10750
CAGTGCCTGG	CTGAAGGCAG	TGGAACAGGG	CCAGCTGGA	TGTTTCTCT	10800
CTGAGGAAGT	TCCTCATCTT	GGTTTTAGGG	CCATACCTTG	TGACCTGTGA	10850
GCTAGGGGTT	GCCAGTCCCT	GACATTCTTA	CTGAGGACTC	GCCTGTCTAT	10900
ATTCCCGGCG	TGTATGTGTC	TCTGAGGTT	CAGACACAC	GGGCGAAGCG	10950
CCTGATGGAT	GGAAAGTATG	TTTTGGTGT	TCCATTGGA	TCTCAAATT	11000
TACAAAACCTT	AGTGCCCCCT	CTCCTCCCTG	TTCCCTCCCA	TCTTCAGTCT	11050
ATCACCTGTT	CCTCATCCAG	CAAATGATAT	TACCATCTTC	CAAGGAGCTT	11100
CCCAAGGAGTA	ATCCCTTGAC	CCTCCCTAAC	ATCCAATTAA	TAATCAAATC	11150
TAGGCCAGGT	ACAATAGCTC	ACGCCCTATAA	TCCCAGCCT	TTGGGAGGCT	11200
GAGGCAGGTG	GATCATTGGA	GGCCAGGAGT	TCAAGACCG	CCTGGCCAAC	11250
ARGGTGAAAC	CTGTCTCATT	AAAAAAAAGT	TATTTTAAA	ACTCAAATCT	11300
ATTATTTCTA	CCTCTAACTG	TGTCTGAAAT	TTATCCATCT	CTCTCCATCT	11350
CTGAGCTGTT	ACCTTACCTAC	AGTCTCAC	GTTCAGTCTA	CGTTAACATG	11400
ACCCAGAGTCT	TGTTCTTAGT	CTGGTGAGGT	CACTCCAGCT	GCTTCAGATC	11450
CTTCCATGGC	TCACCCTGTC	CCTCATATAA	AGTTGGACT	CCTGGACATG	11500
TGGCTTACCG	GGCCCTCCGT	GATGTGGCCC	TATTTGCTTC	TCCATTCTGT	11550
TCTCTCCCG	CCTCTCTGCC	CCCATCTCTA	GGCACCAAC	ACACCCCTCT	11600
GCTCGTCAAT	GGTGCAGCT	TCTCTCTAT	CTCTGGTCTT	TGGACAGACT	11650
TTTCCCTTCA	CCTGGAAATGC	TTTCTTCAT	CCTACCCAC	TCTCTTTAAT	11700
CTAGATAAGG	TTTATTCTTT	TTGAATGTCT	AGCAGTGAA	CCATTTCCTCC	11750
TGAAAACCTT	TCTCTAAACCA	ACCCCTTAC	CTCAGCCCCA	GGCTAGATT	11800
AGGAGTCCCT	CTGAATGTTT	CCATAGCATT	TTAAAGAAAT	TGCCCTATT	11850
CTTGTTCGTA	TCTATCACTA	AACTACAAT	TGTATGAGAA	CAGCCACTAT	11900
CTCTGCCCTGG	TTCACCATTC	ATCTCCAGCA	ACTAGCATAA	TGCCTGGCAG	11950
AGTCAGCTG	CAACAAATAT	TTGTTGAAATA	AATTAACAG	TGGCTTTATC	12000
TCTCTTAAGTA	AATCTGCTT	TTTTCACCTA	TTAAAACAGA	CGCACAGGCC	12050
AGGTGTGGTG	GCCCCATCT	GTAATCCAC	CACTTTGGCA	GGCTGAGGTG	12100
GGCGGATCAC	CTGAGGTCAAG	GAGTTCAAGA	CCAGCCTGGC	CAACATGGTG	12150
AAACCCCATC	TCTAATAAAA	ATACAAAAAT	TAGCTGGCA	TGGTGGTGGG	12200
TGGTATAGT	CCCAGCTACT	AGGGAGGCTG	AGGCAAGAG	ATCGCTTGAA	12250
CCCAGGGC	AGAGGTGGCA	GTGAGCCGAG	ATCATGCCAC	TGTACTCCAG	12300
CCTGGATGAC	AGAGACCTGT	TCTCAAAACA	CACACACACA	CACACACACA	12350
CACACACACA	CACACACACA	CACACACACC	AAGTTGTATA	ATTAAAATA	12400
TAACGTGCTT	GTATGGAAC	ACTTGTAAA	TACAGGAAG	TAATAAAAAA	12450
GTCCTACCATC	TAGCTCACCA	CATAATGACC	ATTGCTATCA	TCCTGGCATA	12500
ATTCTCTCCT	GTATATAAT	ATATTTCTT	TTATTGTTAA	AATTACACTA	12550
TGAGTACTAT	TTATTATTATT	TACTGTGGCA	AAATGCGAA	AACATAAAAT	12600
CTTGCCTATT	TAAGGTATGC	AGTTTGGTGC	ATTCACACCA	CTCACATTGT	12650
TGTGCAATA	TCACTCACT	CTATCTCAGA	ACTTCTCGT	CTTCCCAAC	12700
TGAAAACCTG	TACCCATTAA	ACAATAGTGC	ATCCTCTGTT	TTCCCCCTCC	12750
TACAATTAT	TTTATTGTTG	GTGGTACCA	AACTGAAAT	AGCTGCTTCT	12800
TCCTTACTTA	GTTCAGATTA	GCATTTCCAT	TTATTGTTAGCC	TGTTTTGTA	12850
GGATGCCATG	ACAGATGCCA	TCTTCCCTAG	AGCTCTTGG	GGCTGTCAGG	12900
TATTTTCAGTC	AGGGTGAATT	CGGGTTGATA	ACATTTAAA	ATCTCACTT	12950
ATTCTGAGGT	TCTCTAGTGT	AGAGCCAC	GTATTTTAG	GGACTCCCAA	13000
GTTACAAACA	AAAATATGGT	GAGGAGGAAT	CACTGAAGTT	TTAACACAAAG	13050
AGACTTACAT	TTGTTCAAT	TTCTATCTT	TAGTTTATT	CCTAACGCATA	13100
AAGAAATACT	TTGAAAATT	TACATAGCAT	TATACATATT	TAATTAAGCA	13150
TGAGCACATC	TTAAAACCTT	AAATTITAGA	TCAGATCTT	AATTCTTAGG	13200
ATATTAAGAG	GTACTGGCA	TTGGCCAGG	TGTGGTGGTT	CACGCCATA	13250
ATCCCAACAC	TTTGGGAGGG	TGAAGTGGCC	GAATTGCTAG	AGCCAGGAG	13300
GTGGAGGGCTG	CAATGGCCTG	AGATCACCC	ATCGTACTCC	AGCCCTGGATG	13350
ATGAGAATGA	AATCTGCT	CAAAACAAA	AAAAAA	AAAAGAAGAA	13400
GAAGAAGTAT	TGGCAATCAG	TGCTCCAGGA	ATAATTCCT	GACTTGAAAT	13450
AAACCTACAT	GTAGACAAAC	TAATTAGGCC	ATTCCAAGAG	TTGCTAGCAT	13500
TGTTTAATA	TGTTTCA	GCATTTCCAGG	AAGCAGTGTG	GCCAGCATTG	13550
CATTTTGTAT	ACTTCAGAA	TGTATGACAG	GTGTTTCTCT	TACCCAGGTC	13600
TTCTGTTTC	TTAGTTTGC	TCATGTAAAT	ATTTATGAAC	ATCCTCATCT	13650

TTTTGAGGGA	AGGGATTATA	GATCATTCTA	ATTCATTCTT	CTAGCATTTG	13700
GTACCATTC	AAGCACATGA	TAGGCACCCA	TTGGAGCAT	TTTGGCTTG	13750
ACAGAATATG	CATTTAGAAT	TGTCAAATT	AGAGGTGTCA	GTGATGGGAA	13800
TTAGAATACT	ATATAATTCT	AAGTCATTG	ACTTAATAC	AAAGAAATGA	13850
TTTCCCTGG	TGGGAATGG	TGAAGGGAGG	CAGGAGTTAA	GAAGAGGAGA	13900
AGAGATCCTA	AGTCATTAT	AAACTCTCT	GGAAAGACAG	GTGTGTGAAG	13950
ACTTTTTAAA	AAAGTCATTCA	CCAAATTGTG	TGTGTGTGTG	TGTGTGTGTT	14000
TTAAATAGAC	TTTATTTTT	AGAGCAGTTT	TAGGTCACA	GCAAATTGAA	14050
ATGCAAGGAC	AGAGATTTC	CATAAACCC	CTGGCCACAC	ACATGCATAG	14100
CCTCCCTCAT	TATCAACATC	CCCACCAAGAG	AGGTGTTTGT	TCTAGTTGAT	14150
GAACCTACAC	TGACACATCA	TTATCACCCA	AAAGTCATAG	TTCACGGCAG	14200
GGTTCACTGT	CGGTGTACAT	TCTATGGTT	TGACCAAATG	TATAATGACA	14250
TGATATCCACC	ATTATAGTAA	CATAAGAGT	ATTTCAGT	CCCTGCAAAT	14300
CCCCTGTTCT	CCACCTATTTC	ATCCCCTCC	CTCTGCATTT	CCACCCCCAG	14350
CCCCTGGTAA	CCGCTGATCT	TTTACTGTC	CCATAGTTTC	GGACGATCTA	14400
TTTTTCAGAC	AGACACAGAG	CTGTCCTTCC	CTTAGTTTCT	ATTCTATCAT	14450
TTCTTCTCC	CCATCCATCA	TTAAAGGCTA	TGAGTTTTT	TTAAGTGTG	14500
AAACACATCC	TACTTGTCA	TTAAAAAAT	AAGCTCCTGG	CTGGGTACAG	14550
TGGCTCATGC	CTGTAATCTC	AGCATTGTTG	GAGGCTGTGG	CAGAAGCATC	14600
ACTTGAAGCC	AGAAGTTGA	GACCAGCCTG	GGCACACATAG	CAAGACCCCA	14650
TCCCTCCACA	CACAAACACA	CACACACACA	CACACACACA	CACACACACA	14700
CACACACACA	CACAAACACA	AGCTCTTGCC	AGAATTAGAG	CTACAAATTG	14750
CCCTCAGGTT	CCTAGAAGAT	AGTCCTTCA	ATTAGATTCA	GATTGAGATG	14800
CTTCCTCTT	AAACAAATGA	TTCCCTTCT	ATCATGCCCA	ATAAGAAAAC	14850
AAATAAAAAT	AAACAAATAC	TGCTGTAAAT	CTCAGCTACC	CAGGAGGAG	14900
AAGCAGAACT	GCTTCAACCC	GGCAAGCAGA	AGTTGCAGTG	AAAGTGAGATC	14950
GCGCCACTG	ACTCCAGCT	GGGAAACAGA	GCAAGATTCT	GTCTCAAAAA	15000
CAAAACATG	TGATTTCTC	CTCTAACTGC	TGCAAGGGG	AATGTTAAGA	15050
ATAGGTC	CCGAAAGAAGA	AGGAAGTAAG	AATGTTTGAC	TAGATTGTCT	15100
TGGA	AAAGTATAC	TGCTGTTGT	CTTCCTAAAC	GTTCCTCAAA	15150
GCTTCG	TTGGCCAGAG	GCTGTCCTCC	TGCGTACCTG	AGGTTTGTG	15200
GCACCAAGAC	AGACTTCCTA	ATTTTCGATC	CCAAGAAGGA	ATCAACCTTT	15250
GAAGAGAGAA	TTTACTGGCA	ATCTCAAGTC	AACCAGGGT	AAAATTTTA	15300
AAAGATTCACT	CTATATTTA	ATTAACGTCA	GTCCGTCTG	AGAATGCTT	15350
GAGAAAATG	TTATTCCTCA	CACCTAACAA	TTAATGAGAT	TAACCTCTC	15400
TCCCTCTCATC	TGACCTGTTG	AGGAATCTGA	ACAAGAGGAG	GAGGCACTGG	15450
GCAGGTTTCC	TTATCATGAT	GTTTGTCTG	TTCACTGTG	GGCCTCACAA	15500
AAAAAAA	AAAAAAA	GGCGTCTG	ATATAACTGA	GAGGTCATTG	15550
TACAGTAAAT	ATTAATAAA	AGCTGATTGT	AGCTGAAGAG	TAGAAGTGT	15600
TGGAGGAGC	AACTGGGTAG	AATCGCGTCA	AACTAAAGAG	CATTCTAGC	15650
CAAAGACACA	ATGATAGATT	GAAGGATATT	TATTCTAAAT	ATAGAATATG	15700
GGTGAACGAG	ATCTGTGGAC	TTCTGGGCTC	CAACGTTAGA	TTCTGATT	15750
AGCAAGCTTG	TCAGGGGATT	CTGATATTGA	AAGGCTGTGG	CCTTCACCTG	15800
AGAAACCTGC	CCTAGGGGGC	CATGAAAATT	TGTCCTGTCT	TTCAAGGTG	15850
CTATCAGACA	TCAAATGGAA	GTAAATCGT	ATCTAAACAA	TTACTAGGAT	15900
GGGGCGACTG	ACTTCACACT	GTAATCCAA	CACTTTGGGA	GGCTGAGGCA	15950
GGAGGATCAC	TTGAGCCCCAG	GAGTTGGGGA	CCAGGCTGGG	CAACATAGAG	16000
AGACGTTGTC	TCTATTTTT	AAATAATTAA	AGAGAAAAAA	ATACTGAAAAA	16050
TATTGTATAC	ACCACTGAAT	TATAATAATG	TGTATATAAT	GTATATATT	16100
ATTATGAGGA	ATATTGATT	ATTTCTATA	TTATATCTTT	TCCTCTGTG	16150
TATTTTATCC	AGTTATGAA	TATTTGAAAC	AATTCATCG	TAATTGGGGC	16200
TAAATTGACA	GAATAGTAAT	CAGAGAAAAT	AGAAAAAGAC	AGATGGGTTA	16250
TCTTGTAAATA	CCAGGTTGGA	GTTGTTATG	GGTTTGTGTT	TTGTTTGTG	16300
GGGTTTTT	TAGACAGAGT	CCCACCTCTG	TGCCCAGGCT	GGAGTGCAGT	16350
GGCACAAGCA	TGGGATAGCA	CATCTGTGAC	CTCTGGGCT	CAAGCAATCT	16400
TCCCACCTTA	GCCTCCTGAG	TAGCTGGGAC	CACAGCTGCA	TGTCACCA	16450
CCCAAGCTAAT	TTTTTTATT	TTTGTAGAGA	CACTTTCT	ATGTTATCCA	16500
GGCTGATCTC	AAACTCCCTG	ACTCAAGTGA	TCCCCCTGCC	TTGGCGTCCC	16550
AAAGTATTGG	GATTATAGC	ATAGCCACCA	CACCCCAACCT	AGTTTCTATT	16600
TAGACTTGGC	CCTTTCCAC	CAGTCATTG	TGTCCAAAAG	ATCTCAAAA	16650
TGTAGACAGG	AAACTGTCT	TTGCTCATCA	TTTTCTTCA	TCCTGTGTCT	16700
AGGGGGATGG	TCGGTGGGGG	AAACTGGGGT	TATGCAAGTT	CCTCTGAAAC	16750
ATCTCTGTG	AGCCCAGGG	TGGATGAGGC	ACCAGCCGCC	AGGGAGTCAG	16800
TGTGCACTG	TCCAGAAAGG	AAGTCATCG	CCAGTCAGCC	GGCCCTGGCA	16850
GCCAGCACCC	GGCAACCTG	CTGCTTGTG	ATAAGAAAT	GGTCTGCC	16900
ACAGGATGGT	GTGGATTTT	CTTTTTCTT	TTTTTTTTT	TTGAGACAGG	16950
GTCGGCTCT	GTCGCCAGG	CTGGAGTGC	ATGGGGGAT	CTGGCTCAC	17000
TGCAGCCTCT	GCCTCCCAGG	CTCAAGGCAT	CCTCCACCT	CGGTCTCCCG	17050
AGTAGCTGGG	ACCACAGGCA	CACACCACCA	CGCCCAACTA	AGTTTCTGTA	17100
TTTTTAGTAG	AGGCAGGGTT	TTACTATGTT	GTCCAGGCTA	GTCTCAAAC	17150
CCTGAGCTCA	AGCTATCCAT	CTGCTTGGC	CTCCCCAAGA	GCTGGAATTA	17200
CAAGCGTGA	CCACTGTC	TGACCAAGGG	GGATTTTTT	AAGTGCACAT	17250
GTGTGGTGTCC	CAGAAGCTCT	GATGGTACCA	AATTCCAAGC	GAAAAAAAGT	17300
CAATGGTTC	CACCCATCT	ACCTCCCAG	ATGGCAAGAG	GAATATCCA	17350
CACTGCAGAT	ACAGTCCATG	TTAAACAAAT	TGCTATGGAT	TTGAAAGTG	17400

AACCTTAAGA	GAAC TGCACT	ATGTTTCTT	CATTAGAGTT	CTCTGGTAAT	17450
TTCCAGCTT	TTTTTTTTT	TTTTTAGAC	AGTGTCCTCGC	TTTGTCGCC	17500
AGTGTACCC	AGGCTGGAGT	GCAGTGACGT	GATCTCGGCT	CACTGCAACC	17550
TCCGCCTCGT	GGGTTGAAGT	GATTCTCTG	CCTCAGCCTC	CTGAGTAGCT	17600
GTATTTAGT	AGAGACGAGG	TTTCACCATT	TGGCCAGGCT	GGTCTCGAAC	17650
TCCTGACCTC	AAAGTGATTG	CCCACATCTCG	CCTCCCAAAG	TGCTGGGATT	17700
ACAGGTGTGA	GCCACTGAC	CCGGCCAGTA	ATTCAAGCT	TCTGAGGAGC	17750
CCTTGAATT	GTTAAATAAC	TTGAGCTAT	GTCCAACATA	TCCATGTTCA	17800
GTGTATGTT	GATATTCTT	AGGAAACCTG	CCCTTGGTTG	TTTTCTTGT	17850
GGTAATTCTAT	GAGCGGCAA	ATTTGACATG	TGTTACAGAA	TATACCTTTT	17900
CTCTGCTCTC	CTACACTCA	ACCGAAACTT	AATTATCTG	CTTGTAGTCAC	17950
ATAAAATAGCT	AACTAAATA	ATATATGAGA	TTTCAGTCTG	CTCACTGTGA	18000
AAATAGACCT	TCTAAATGAT	CTCTTCCACT	TGCAGATATT	TGCAAATATG	18050
GATCCATCCC	TCCTGATGTG	GAGGAGAAGT	TACGGTTGGA	ATGGCCCTAC	18100
CAGGAGCAAT	TGCTACTCCG	AGAACACTAC	CAGAAAAAGT	TCAAGAACAG	18150
CACCTACTCA	AGTAAGAAGA	AAAAGGCACC	CTAGAGATGT	TCAGCCCCA	18200
AAGATATTG	AAATGGTTGG	ACTCGGGCACC	CAACTCTAGC	AGTCCTACGG	18250
AAGTTGTATA	AAGCTGAAAAA	TACTGAAGCA	TTTCCCAAAT	GGGAAATCCT	18300
AAACTCAAAA	CTTGTCTTTT	GGTTTTTTG	TTTGTGTTGTT	TTTTCTTCAT	18350
CTGACATTGC	TTAGTAGTCA	CAGAATGAAA	GATAATCAA	TCATTGATGA	18400
TCTAACATG	ACCTTCAGTG	CTCTAAAAAA	CTACGGAGTC	AAGGAAACAA	18450
TGAATATATT	CCTCATGTAA	AATTAAATA	CAGACATATA	AAGGGCAAAA	18500
CATGAAACATC	ATTCTACACT	TGAGGTTCCGT	CCCCCTCCCA	GAATAAACCC	18550
CCAGTATGCC	TTGGTTAGA	GCATTAAGCA	GGAGGGCCCT	GAGTCACTCC	18600
AGACAGCTT	GACCAACCAAG	CAGCATTC	TTTTGTTC	CTCTGGGCT	18650
TTTGCACAAACA	CAGGGCTAGC	TCAGCTACCC	ATTAGTATGT	TTTCAGTCAC	18700
AAAAAACAGTC	TTCCAGTCTT	CAAATTAGGA	TGACATTGTC	ACATGGGGCT	18750
TTAAAGCAAG	TGAAACAAGG	AACCCCTTTT	TTTTTTTTT	TTGAGATGGA	18800
ATCTCACTCT	TGTCGCCAG	CCTGGAGTGC	AATGGCGCAA	TCTGGCTCA	18850
CTGCAACCTC	CACCTCCCG	GTTCAAGAGA	TTCTCCTGCC	TTAGCCTCCT	18900
ATTCAATTATG	AGGAATATT	GATTATTCA	TTCTGTAGG	GTAAGAGATAT	18950
TACCCCCGAT	CATATTATTG	ATTATTGAGT	AGCTGAGAT	ACAGGTGCCT	19000
GCCACCACGA	CCGGCTAATT	TTTGTATTT	TTTAGTAGAG	ACAGGGTTTC	19050
ACCATGTTGG	CCAGGCTCCA	GGCTCGTCTC	GAACCTCTGA	CCTCAGGTGA	19100
TCCACCCACC	TCAGCCTCCC	AAAGTTCTG	GATTACAGGC	GTGAGGCCACC	19150
ACTCTGGCC	ACAATCCTT	TTTAACATG	AAATATATT	TTATCTGAAG	19200
TTTGATGTTT	ATACCTTCA	GAGGGATGAT	GTTCCCATAT	CTCAGTTAAA	19250
GAATAAACCT	GCTCAGATAC	TCACAGCTCT	TCTTTGACT	TTTGAAAATA	19300
AATGATCTG	AAGTTACTAT	ACTTTGTTG	GGTTAGTAA	CATTATTTAA	19350
AGTATATTAT	TTTAATTAA	TATCTTTGTA	AGATTTTACT	GTATACTACC	19400
TGGAGTTCAA	TGTATCAGAT	GGTTTCAAA	TTTATGTA	TTTTTATGT	19450
ATATGGTACA	AAAAAAAT	TGATCCATAA	GAAATCAGAA	AAATAGCGCAT	19500
ATGCTAATAG	CTAATGTTG	CCTCTAAAAA	ACTTATTTT	GCATTTTTAA	19550
GAGGGGGATA	TACTCTGACA	CTTTAATAAG	TGTAATTAAT	TATTGACTGG	19600
AATTGGCAT	GAGGCAGGGC	CATTTCA	CCACTTAAAG	GAATGACACA	19650
TACCAAGAGA	CCACACCAAT	TGTAATAAA	TCATTATAGC	19700	
TCTGCTAGGA	GAAGACCCAG	TTGTTAGG	TAATTATGG	ATTGCTCTT	19750
AAAACACATG	TCCCAGAAGA	TATAGGTGAG	TCTTGGGGGG	CCGCATTAAA	19800
CATTATACCA	ATGTATCTTA	CATTCTCA	AAAGTTTAC	TACTTTACAG	19850
GATCTTCTG	TTACCAAAAT	GGAAAGGTTTC	CAACTCCAGG	ACTTGGCTTT	19900
CATAGTTCT	ACACCAAGGG	AAATGCCCTC	TTTGTCAAC	TATGCAACCA	19950
GGTTAGTTAG	TGTAAGTCCA	GCCACCTGT	TGGCAATGCT	AAAAGGTACA	20000
ACAAACACAG	AAATTTTTT	GCATTGTA	ACATTGATT	TCTGGCTCGA	20050
AATTTCAGT	TTTCATGGGC	ACGTCATGGA	AACAGAAATC	TTCTGTGTTT	20100
AGTTTGGCA	CCTACTCTT	GTAATGACAA	ATATTCA	AGCCAATAGG	20150
GGATTCCACA	AATGTTCTG	AACCTGTTG	TGAGACTGGT	AATGGCTGAG	20200
TGACATGGG	ACATACCA	AAAGAAGAGG	TAGCAAAGG	CTGCTGAGAT	20250
AGGACATGT	TCATTGCTTA	GCTAGTGGCC	TGCACCCCTTA	AAACACATGT	20300
CCCAAGCTGG	GTGCTGTGGC	TCACGCCCTG	AATCCCAGCA	CTTGGGAGG	20350
CTGAGGGCGG	TGGATTACCT	GAGGTCAGGA	CTTCGAGACC	AACTGGCCA	20400
ACATAGTGA	ACCTCATTC	TACTAAAAAT	ACAAAAATT	GCCAGGCATG	20450
GTGGCGGGC	CCTCTAGTC	CACTACTCA	GGAGGGCAGGC	AGGAGAATTA	20500
CTTGAATCTG	GGAGGAGAG	GTTGTTGTTG	GCCGAGATTG	CGCACCGCA	20550
CGCTAGCTG	GGCGACAAAG	TGAGACTCTG	TCTCAAAAAA	ACAAAAACAA	20600
AAAACAAACA	AAACAAAAAC	AAACACAA	AAAAAACGGG	TATCCCAGAA	20650
GATACAGGTA	AGTTTCTAA	CACAGGTCT	CTTGTATGTT	GCCTTCCACT	20700
TAAGTAGAAG	ATGACAAAAA	CATTGTCAT	GAGAATATAG	ACTCACATT	20750
TAAACCTGTT	TGAGCAGGAA	AGGAAGGAA	TGTTACAGAT	GTAATTCTGG	20800
GTGTGACTGC	AGAAAGGATG	ACTCCCTAT	TAAAGTAGTC	ATCTGAGTG	20850
AGCTTAACCTCT	TTGTTACTTC	TCTTCTCTC	CTGTTCCCT	CATCACCCCA	20900
TTCTTCCGTT	GCCTACACCC	AGGCCCCAC	TGGATGCTGA	CATAGACTTA	20950
CATGGTACAG	TCCAAGGGAA	AGATCTGCCA	TTTTTTCAA	TGTGTCACT	21000
TGGTTATCTT	CATTCCAAGG	ATCTCTCCAC	TCTTTATACA	GTAAGAGATG	21050
AGAGTCTGGA	AAGGATTGGG	AATAAGATAA	TGAATTGTA	GTTTAAATT	21100
GTTCTTCGTA	TTTGGGGAA	GGAGTAGGCT	AGGTGGCCT	TCTGTTTTT	21150

TTTTGTTTTT	TTTTTTAAAG	TAGATGTGGC	CAGACGTGGT	GGCTCACGCC	21200
TGTAATCCCA	GCACTTTGAG	AGGCTGAGG	AGGTGGATCA	CTTGATGTCA	21250
GGAGTTCAAG	ACCAGCCTGG	CCAACACAGT	GAACCCCGT	CTTACTAAA	21300
AATACAAAAA	CTAGCCGGC	TTGGTGGCGT	CCACCTGTAG	TCCCAGCTAC	21350
TGCAGAGGTG	GAGGCAGGAG	AATCATGCA	TTGTACTCCA	GCCTGGCGA	21400
AGTGAGCCAA	GATCATGCCA	TTGTACTCCA	ACCCGGAGG	TGGAGGTTGC	21450
CTCTGTCTCA	AAAAAAAAGA	GAAGAAAAAA	GAAGAAAAAGA	ATGGATTTGA	21500
ACTCAGTCGT	CAATAGCCTC	TATTCCAGGA	GATGTTACAG	TTGATTATGT	21550
TATAGGGGT	GTATAATAGA	ATTTCCAGCT	ATGTAATTTC	CAAGTGCATT	21600
TGGAAGAATG	AAGAAATGGA	GGAAAGGGTA	AGTATGAGTG	CAAGCATTC	21650
AGGTTTTTTG	AAAATGCTAT	AATCTTGTG	CAGGCTAGT	ACAAAGTGCT	21700
ATTTAGCTGT	AAAGGTTTT	TGTGATTTAC	AGACAGTTT	CACATGTGTC	21750
ATTCACACTT	TGGTTTTATG	GCGAAGGCAT	GTGATGGTGC	TTGTCCCAGG	21800
ACTTATGAGC	CATATCTGAG	GTTCTGTGTC	GGCAAAAGA	TTACCCCTGA	21850
TCATATTATA	GTCTATAAGT	GGGAGAGTTG	TGCCCTGGAGC	TCAAGTCTTA	21900
TGATTTCTGA	TCCAGGCCAC	TTCCCTAAC	ATGATTTTG	AAATAAAAAG	21950
CCTATAATGT	GTGACTAAAG	CAGGTCACTC	ACCCCTGTG	ACAGACTCTA	22000
GTAATGGTAC	TGCAACCAAA	CGGCTGCGTG	ATATTGGCA	AAAGACTTAC	22050
TTATTTGAAT	CTCAGTTCC	TCCCTAGAAA	ATGAGGGTGG	AGGTTAAC	22100
TCAGGCTGATG	ATCCTAAAGC	CTCCATCATG	CCCTAAACTG	TGGCTCTAAC	22150
ATCCAGTAGA	ATGCTGGTC	ACAGGACTCT	AGGGAGCTT	TCAAACCCAA	22200
ATGTCCTGTC	TTCCCTTGATG	GTAGGCAGCA	GTTTATGGAA	GTGGGGCGACA	22250
CAGCAAAATAT	CAAATAC	AAAGCAGCTT	GCAAGAGTTG	TTCTGCGCTA	22300
GTGGTCTTTA	TAGTTAATAT	TAAGTATTA	ATTTTTTTT	TTTTTGAGAC	22350
AGAGCTTGC	TCTGTTACCC	AGGCTGAGT	GCAGTGGCAC	AACTCTGGCT	22400
CACTGCAACC	TCCACCTCCC	GGGTTTGAGC	AATTCTGTCT	CAGCCTCCCA	22450
AGTAGCTGGG	ACTACAGGTG	CATGCCACTG	CACCCAGCTA	ATTTTTGTAT	22500
TTTTAGTAGA	GACGGGGTTT	CACCATATTG	GGCAGGCTGG	TCTCGAACTC	22550
TTGACCTCTAG	GTGATCCACC	TGCCCTCAGCC	TCCCAAAGTG	CTGGGATTAC	22600
AGGCATGAGC	CACTGCACCC	AGCTTAAATA	GCTAATATTT	AAATTATTTC	22650
TATAGTTATT	CAAGTAATTG	AGGCCAAAGA	CTTAGAAACA	AAACAAAAAG	22700
CCACTTTAA	GGAGAAAGGG	TGTAAGTTG	CCAGATAGAT	AGAGATCTT	22750
CTTTTTAAC	TACAAGAGTT	CAGGAATGAA	TTACTCTTTA	ACAAACGACT	22800
ATAGATATAC	ATGAAAATTG	GAAGGACTTA	TTATGCAAT	GATAATCAAT	22850
TTAAAGACAA	CACTTAAAAT	TATATTGTTG	CCACTCTCAA	AAAGTGGTAA	22900
TAGAACAGCT	ATAGGTTAA	AAAGCAGAGT	ACAGAACGTT	CCAAACTTAT	22950
GGCACCTTAA	TATCGCAGAA	AACCTTTAA	AGCATGCTA	GGCCACAAAAA	23000
AATACCTGTA	TTTTGATTAT	AAATTGTTA	GGTCTACACA	ACCTAATAGT	23050
AATAGGTCCTA	ATAGTAATGC	TGTCCTATAG	ATGTTGATGT	TTTTTCCCT	23100
GCAAACCTAA	AAGATCTAC	AGTGCCTCTG	TAATAGCAC	TGCCCTGGTTA	23150
GAGTTGAAAT	TCAGATAAAT	TTTTTTTTC	ATGTTAATTA	TTTTTCTTTT	23200
CTTTACTTTT	TTTTGTTTTT	TTGTTTTTT	TTTGAGACA	23250	
GGGTCTCAT	CTGTTGCCA	GGCTGCTGTG	CAATGGCATG	ATCATGGCTC	23300
ACTGCAGCCT	TGACCTCTCC	GGGCTCAGGT	GATCCCTCCA	CCTCAGCCTC	23350
CCAAGTAGCT	AGCTGGGACT	ACAGGTGCTT	ACCATCATGC	CCGGCTAATT	23400
TTTGTTTTT	TTGTAGAGAT	GTGGTTTGTG	CATGGTGGCCC	AGGCTGGTCT	23450
TGAACCTCTG	GGCTCAAGTG	ATCCGCCCGC	CTCGGCCCTCC	CAAAGTGCTA	23500
GGATGACAGG	CATGAGCCAC	TGCACCTGGC	CCCTGGGGGA	AGTATTCTT	23550
AATGGTTACA	TAGGACATAC	ACTAAACATT	ATTATTTGTC	TATATGAAGT	23600
TCAAGTTAA	CTAGGTGCCG	TGCACTTTTA	GTTGCTAAAT	CCTGTAGCTG	23650
TACCCATGCA	TTCACTGGTG	CTCCCCAGCT	TGCCTGAC	AGAGTTGG	23700
AACCATAGTC	CTATAACTCT	AGGCCAAATT	TTTAATGTA	AAATTGATTC	23750
ATTTTAAATT	AATAAATAAT	AAACAGGAATT	TTTTTAAATAA	TTGTTTTAAA	23800
TATAATTAAA	ATTATCAAA	TATTTTTTAA	CTGAATTTGT	GACTAGAGAT	23850
ATTTAGATTA	TGAAGAGTGG	GGTTTATGCT	AACTAATGAC	AGCTTGGCTA	23900
TGCTATGTGGA	GCACTGAGCT	ATAAATTG	GCTTCCCAA	TTCTCCTGAT	23950
GTCACTTGTAA	CAAACCTAA	GTGTCAGAC	AGAGCTCTG	GTATCTTCCA	24000
TGGGATTTC	TCACAGCT	GGAGCAAATG	AACTCAGATT	GATTTTTTTT	24050
AATTGTC	ATTTGTTGT	CTCAAAAACA	TAATTATAAT	CATTATTAG	24100
AACTAGAATT	TCTTCAGTTT	AAACACAGAA	ATAGTTATTC	ATTATGAAA	24150
GGCAATCTGG	AGGCCTTCAT	TGTGGTCCCA	ATCTAACCT	TAATTGTGA	24200
CGTTTTCTT	TTAGGAAGCT	CTGCTAGATG	GCTATACACT	TTTGCAAACT	24250
GCTCAGGACT	GGACTTGTAC	TTGGCTCTAA	ATGCTTATT	AAAGAACAGCA	24300
GATTTGCA	GGAAACAGTTC	TAATGCTAG	TTGCTCTGG	ACTACTGCTC	24350
TTCCAAGGGG	TATAACATT	CTTGGGAAC	AGGCATGGT	GAGTACCCCA	24400
GGGAACAAATT	CATTAATAAG	GAGATTCCCC	ACTAGCATT	TTCTTTTCT	24450
TTTCTTTTTT	TTTTTTTTT	GAGACAGAGT	CTGGCACTG	24500	
TGCCCAAGGCT	GGAGTGCA	GGGCCACCT	CGGCTCACTT	GAAGCTCTGC	24550
CTCCCAAAAC	GCCATTCTCC	TGCCCTAGCC	TCCCGAGTAG	CTGGGACTAC	24600
AGGCACCCGC	CACCGCGCCC	GGCTAATT	TTTTTTTTT	TTTTTTTTT	24650
TTTTTTGCA	TTTTTAGTAG	AGACGGGGTT	TCACCGTGTG	AGCCAGGATG	24700
GTCTTGATCT	CCTGACCTCG	TGATCTGCC	TCCCTGGCCT	CCCAAAGTG	24750
TGGGATTACA	GGCGTGGAGCC	ACCAGGCCCG	GCTAGCATT	TTCTTATGA	24800
CACTTTTTT	TTTTTTTGTG	GACGGAGTCT	CGCTCTGTG	CCCAAGGCTG	24850
AGTGCAGTGG	CGCCATCTCG	GCTCACTGCA	AGCTCCACCT	CCCAAGGTTCA	24900

CGCCATTCTC	CTGCCCTCAGC	CTCCCGAGTA	GCTGGGACTA	CAAGCACCCG	24950
CCACCAACGCC	CGGCTAATT	TTTGATTTT	TTAGTAGAGA	CGGGGTTTCA	25000
CCGTGTTAGC	CAGGATGGTC	TCTATATCCT	GACCCCATGA	TCTGCCCGCC	25050
TCGGCCTCCC	AAAGTGGTGG	GATTACAGGC	GTGAGCCACT	GGCCCCGGCC	25100
AACACTCTTT	TTATTATTAG	CAAATATACT	TCTGCCCTGG	CACATTCTTG	25150
CAAGTGTCA	ACAATGCAAC	TTTGGAAGT	GCATGTGGCA	GAAACTCCTG	25200
CTGTATTTAT	TCCAGAACCT	ATTATTGCTA	ATCCCAGTT	ATGTTACATT	25250
TGAAGTGTAGA	ACCAGTTGGA	GCCAGCAACG	TTCCCAGCTC	CAAAGTTCCC	25300
TTGAGATTTT	CAGAATCACT	TAACCCATT	ATGCTTGGA	ACCTGGACTC	25350
AGCAAAACTG	GGAAAGTCAGC	AGTTGTTTT	ATTACATCCCT	TCTTTCTCA	25400
CTTTCTCAAA	TGTGTCAGTT	AATCTCAGTA	ACCCCATTCG	AACCTTCATT	25450
ACCTGCCAA	GCGGTCTAGA	ACTTGCACT	ATAGAACCT	ACGTTGGTCA	25500
AGCTCTGTAC	TGTCTCTTC	TTCACTCTT	TTTGCAAAAG	AACTTGTAAA	25550
TTTAACAT	AAGTATTCTA	GATTGCGCAC	ATTTATTCAA	AACATAGAGT	25600
GCTTTTCTCA	CATATCAGCC	AATGGAAATA	AGGATAAAT	GGGAATGAA	25650
ATGTAGTAAT	AGGATAAGCA	CAAGTCTCT	TCCGCTCAA	ACTTTTTTTT	25700
TTTTTTTTT	CAGACAAGAT	CTTGCTCTGT	TACCCAGGCT	GGAGTGCAGT	25750
GGCGTGTTC	TAGCTCAATG	TAACCTCCAA	CTCCTGGCT	CATGCAATCT	25800
CTCACACCTC	AGGCCCCCTGA	TTAGCTAGGA	TCACACTATG	CCTAGCCAAT	25850
TTTTTTCTT	TTGTCTGTT	GTGTCGCCA	GGCTGCTCG	ATCTCTGGC	25900
CTCAAGTAAT	CCTCCGCTC	CGGCCCTCTA	AACTGCTGGG	ATTATAGGC	25950
TGAGCCACTG	TGCCCCGGCT	CAAACCTTT	TTTCCAAAGT	AAATGAAGTT	26000
ATTAGATATG	GAATATAGTC	TAGTTCCCG	ATATCCATAT	CCATTGGTTT	26050
ATTACCCCTCA	TTATTAACCT	AAATTGTTT	AATAGACCC	CATATCTCAG	26100
TTATAACAGT	AAAATTGTT	TTTTGTTTT	CTGGAGTATC	TTATTTATAA	26150
CTATGAGTTT	TACTTTACTT	ATTATTTTTA	TTTTTGAGA	CAGACGCTTG	26200
CTCTGTCACT	CAGGCTGGAG	TGCGGTTGCG	TGATCATGGC	TCACATATGGC	26250
CTCGACCTTC	TGGGCTCAAG	TGATCCTCTC	CCTCAGCCTC	CCAAGCTGAG	26300
ACTACAGGCA	TGCAACACCA	CATCTAGCTA	ATTTTTTTT	TCACCATGG	26350
AAACAAGCCTT	TACTATGTTA	CCCAGAGTGG	TCTCAAACCT	CTGGCCTCAG	26400
GGGATCCTCC	TGTCTCAGCC	TACCAAATG	CTGGGATTAC	AGGCATGAGC	26450
CATAGCGCCA	GACCTGGTTT	TACTTTCTCTT	GACTTGTAA	TACAAGTTTT	26500
TGTAATTGG	AAAATGTTT	GTTGCTTTA	AATACTGCTG	TATGTTTGCT	26550
TTAAATACA	ACATTTCTCG	ATATATATT	TGAGAATTGC	TGTCTTTCA	26600
AACCTAACAG	TTTCTTAAG	AAGGCTGATA	TTTCATCAA	TGGTCGCAG	26650
TTAGGAGAAG	ATTTTATTCA	ATTCGATAAA	CTTCGATAAGA	AGTCCACCTT	26700
CAAATAATGCA	AAACTCTATG	GTCCTGATGT	TGGTCAGCCT	CGAAGAAAGA	26750
CGGCTAACAT	GCTGAAGAGG	TAGGAACCTAG	AGGATGAGA	ATCACTTTAC	26800
TTTCTCTCTT	TTTCTTTTG	AGACAGAGTC	TCACTCTGTC	AGCCAGACTG	26850
GAGTGCAGTG	GTACAATCAT	GGCTCACTGC	AACTTGCACC	TCCCAGGCTC	26900
AAGCAATCCT	CCCATCTCAC	TCCACAAAT	AGCTGGACT	ACAGGTGCAC	26950
ATCACACAC	CTGGCTACTT	AAAAAAAT	TTTTTGAGA	GATGGGGTCT	27000
CCCTGTGTTG	CCCAGGCTGG	TCTCTTGAAT	TCCTGTGTC	AAAGCCATCCT	27050
TCCACCTCAG	CCTCCCCAGAG	TGCCAGGATT	ACAGGCATGA	GCCACACAC	27100
CCAGCCACCA	CTTTCTTAA	AAAAAAAGAA	AGATTCTCTC	TGGTAGACAA	27150
TCCTCAATAG	TCACACATGTT	ATTAACACAT	CTGCTGCTG	ATACATGAT	27200
TTACAAAAAA	AAGGAAATT	TGACGGGTTC	AGAAATACAA	GGGATCTGAG	27250
GCAATGTCA	CCTATGATAA	ATTTGCTAT	AAAAATTAGG	AAGTTTGCT	27300
TTACCTGATC	CTAAAGCAAT	AACCAACCA	TTTCTAGGGA	ATAAAACCT	27350
CATCGTATA	TGTTGCAAT	ATATGTATA	TATGACTGAG	TGATAATAAA	27400
ATTTTTTTTC	TAGCTCTCTG	AAGGCTGGTG	GAGAAGTGT	TGATTCAGTT	27450
ACATGGCATE	AGTAAGTATG	TCTCCTATTC	TTAATACACT	GAAAGTAAGG	27500
CTAGCTTTAT	TTATTACCTA	GTATTCAAAA	AGTTAGTCA	TTTAACATGCC	27550
AATTGACTGC	AGTTCACATA	AGAAACAAAT	AGTGTCTCAA	GTAGCACTGT	27600
ACTCCAATT	TAATATTAAT	AAAAAAATT	TTAAGTTATT	TTAAATAATG	27650
TAGGGTTTC	TATAAGATC	ACTTTATACA	GAAGAACAGT	GCCAATTAAC	27700
CCATGGAAACA	TATAAGTAGC	AAAACCAAT	TGCTTGCCTA	AGAACCCAGTA	27750
ACCCAGGAGT	ACATGTCCTC	GCCACTGTGT	TTTTCAAGA	CAGAGTAAC	27800
GATTCTCTAGT	TACTTGCTA	GAATGGACTC	CTCCTCATAA	CTCCCTTCCA	27850
TCTGGCTCTT	TCCCTAGTAG	AACTTCTACC	TTTTTTAGT	AAACAGGTGAG	27900
TGGGAGAGGT	AAGAAGGAGA	ATAAGGTCA	CAATTACCT	AAAAGCAGAA	27950
AGTAAAATT	TTTATTTTTT	TTCCTGAATAT	TTTCTGTGTA	ATTTAGCTAC	28000
TATTGAAATG	GACGGACTGC	TACAGGGAA	GATTTCCTAA	ACCCGTATGT	28050
ATTGGACATT	TTTATTTCTAT	CTGTGCAAA	AGTTTCCAG	GTAATAGTCT	28100
TTTAAACTT	TTTAAATGTAA	AACCAGAAC	CTTATTTTAT	AGTCTAGCTA	28150
GTCTAAATT	CTATAGGTAT	GTATTTTAC	ATGTTTTCT	AATTTAGAG	28200
AACAAGCACT	ATGACTTATC	CACTGTTAGT	TTTCCCTTA	GCATTGGTC	28250
TTACCCCATG	TACGTGATTA	GAATTTGAA	ATATTCCTAA	TAGCCTTTAG	28300
TAGAATTAAC	TCACATAGAT	GATAAGATG	GGTTGGTCA	CTTCATGTT	28350
CTTCCACACG	CTACTATTT	AATAAAAGAA	AGTTTCCAA	GACCTAAATG	28400
ACTATGAACA	TATTTTATAAA	CTATATAGGA	GGGGTGGTC	TAGGAATACA	28450
AAGTTTGAA	TGCTGTTAAT	CTTCACAC	ACAGTGA	CCACAGGTCA	28500
GCTTTTGC	AATTACCATG	GATACTTT	TGTTCTATAG	GTGGTTGAGA	28550
GCACCAAGGCC	TGCAAGAAG	GTCTGGTAG	GAGAAACAG	CTCTGCATAT	28600
GGAGGCCGGAG	CGCCCTTGCT	ATCCGACACC	TTTGCAGCTG	GCTTTATGTG	28650

AGTGAAGCAG	CGCTGGCCTT	AGGGTCAGA	GTGCAGCTCT	TCTCCATCCT	28700
TCTATTCTGC	TGAAATAGCT	CCCCAGCCAA	AAAGCAGATC	AAAGACCGTT	28750
TCAGTGGCTG	AGCCCCAAA	TTCATGCCAG	ATTTGCAAG	AAAATGATT	28800
ACTAAAGCTT	GAGGCCACAT	TTAACAAAGT	GTTCCAAATT	AATCACTATA	28850
AGGATGAATT	TTTCAGAAA	TTTGCCTT	TAATTATGCC	CCATAAAATAT	28900
GTCAAGTAGT	CCTTAACCTA	AAGAAGTACA	CTGTAAGAAGA	ATGCATATAG	28950
CCGGATATGG	TAGTTCCCTG	TAATCCCAAT	ACTTGGGAG	GCCAAGGTGG	29000
GAGGATTGCT	TGAGCCCAGG	AGTTTGAGGC	TGCAGTGTAGT	TATGATGTTG	29050
CCACTGCACT	CTAGACTGGG	CAACAGAGTG	AGACTGTCTT	TTTTTTCC	29100
CTCTGTCAACC	CAGACTGGAG	GGCAGTGGCA	CGATCTCACC	TCACTGCAAC	29150
CTCTGCCTCC	CGGATTGAAG	CGATTCTCCT	GCCTCAGCGT	CCTGAGTAGC	29200
TGGGACTACA	GGAGTATCAC	CGCACTGGG	TAATTTTG	ATTTTTAGTA	29250
GAGACGGGGT	TTTGACATGT	TTCCCAGGT	GGTCTGAAAC	CCATGAGCTC	29300
AAAGTGTCTG	CCTACCTCAG	CCTTCCAAA	TGCTGGGATT	ACGGACATGA	29350
GCTACCACGC	CCGGCCACAC	CCTGTCCTT	AAAAAAA	AAAATGCAAG	29400
TTAGAGCATA	TTACAGCTT	GTCTCTCAGG	AGGATACTTA	GTGTATGTAG	29450
CTATAATTCA	TAGATCCCA	AGAAGTTAG	AGCCTAAAGT	ATGAGGTCCC	29500
ACCAAGGGGG	CTATCATTAA	ATTTAAAGAT	TTGTTAAATC	ATCTCATTGT	29550
CCAACACCAC	AAACTTGATT	GCTTAAAT	ACTGTTTAG	TTACATTAG	29600
TAACTCTATT	AGTGCTTTA	ATCTAATCG	CTATATCCTC	ACATTGAGAT	29650
TTTTTTCTT	TCTCTCTT	TCTTCATTC	TTTTTCTCTC	ATCCTCATTC	29700
TTATAAGCCT	GAATAACATC	ACAAATCCTT	TATGCCCTAG	GAAGCAAGAG	29750
GAATAAAAGAA	TGGAGATGTT	TGTTTGGCA	TTAACAAAG	ATCTGGGGTG	29800
TCGGGGAGAA	GGGGGATAGA	GAAGGAGAAG	TGGGAAGAGG	TGTCCTATAAT	29850
AGCTTAGGTC	CAATTCTGCT	TATTTTACAT	TTTACCCCCG	CTGACTGCCA	29900
CTTTTCTTC	AGGCCCTCACA	CATTTGTTG	GCAGGGACCT	CATAGGACCA	29950
GGAAATTGCT	ATAGAGGTTG	GAATTGTC	CACCCGAA	GGGATACCTC	30000
TAGCATGGTA	ATAGTCTCT	AGGATTGTT	ATCATATGGA	AAAGATGAAA	30050
GGGAGGGATT	CTGCTGCTG	TGCTGCTGCT	GCATGCCAGTT	GCCATTTCAT	30100
TTAAATGACT	TATTTATAAT	GTGACACT	TTTCTGGCTT	CCTGTTAATT	30150
CCTCCCTCAA	AGATAAATAA	ACCPGAACCA	GGCATGGTGG	CATGCACCTG	30200
TGGCCTGTA	ACCAACCAAC	ACGTTCACT	TGCTGCTGCT	CTAGATAGAG	30250
CCAATTATCA	AGACAGGGGA	ATTGCAAAGG	AGAAAGAGTA	ATTATGCAAG	30300
AGCCAGCTGT	GCAGGAGACC	AGAGTTTAT	TATTAACCAA	ATCAGTCTCC	30350
CCGAACATT	GAGGATCAGA	GCTTTTAAGG	ATAATTGCG	CGGTAGGGC	30400
TTAGGAAGTG	GAGAGTGTG	GTGGTCAGG	TTGGAGATGG	AATCACAGGG	30450
AGTGGAAAGTG	AGGTTTCTT	GCTGTCTCT	GTTCCTGGAT	GGGATGGCAG	30500
AACTCGTTGG	GCCAGATTAC	CGGTCTGGT	GGTCTCAAAT	GATCCACCCA	30550
GTCAGGGTC	TGCAAGATAT	CTCAAGCACT	GATCTTAGGT	TTTACAACAG	30600
TGATGTTATC	CCCAAGAACAA	ATTGGGGAG	GTTCAAGCTC	TTGGAGCCAG	30650
AGGCTGCATT	ATCCCTAAAC	CGTAATCTT	AATGTTGTAG	CTAATTGTT	30700
AGTCTGCAA	AGGTAGACTT	GTCCCCAGGC	AAGAAGGGGG	TCTTTCTAGA	30750
AAAGGGCTAT	TATCATTTT	GTTCAGAGT	CAAACCATGA	ACTGAATTTC	30800
TTCCCAAAGT	TAGTTCAGCC	TACACCCAGG	AATGAAGAAG	GACAGCTTAA	30850
AGGTTAGAAG	CAAGATGGAG	TCATGAGGT	CTGATCTCTT	TCACTGTCA	30900
AATTCTCTCA	TTTATAATT	TTGCAAAGGC	GGTTTCAGTC	CCAGCTACTT	30950
GGGAGGCTG	GACAGGAGGA	TTAATGGAGC	CCAGGAGTTT	GAGGTTGCAG	31000
AGAGCTATGA	TCACGCCACT	GCACCTCCAGC	CTGGGTGAC	GAGTGAGACC	31050
CTGTCTCTAA	ATAAAATAAT	AAAGTAATAA	ATAAAATCAT	AAATAAAATC	31100
AAAGATGGTGT	GCATTAGAA	TTGAGCGATT	TTGTTTCCAA	ACCTCAAGAA	31150
AGCTTGGCT	TGCTCTGTC	CAGTGGCTG	GATAAAATTGG	GCCTGTCAGC	31200
CCGAATGGGA	ATAGAAGTGG	TGATGAGGCA	AGTATTCTT	GGAGCAGGAA	31250
ACTACATT	AGTGGATGAA	AACTTCGATC	CTTTACCTGT	AAAGTACCCAT	31300
TATTTCTTA	ATTCTAGTGG	AGTAGATTAA	AGTCAACTCA	GGACCTCTGG	31350
TGTTAACCTC	CTATGAACAG	TCAGTCCTCT	CAGTAACCTAG	CCAAATCATG	31400
AGATGATGAA	TTAGAAGGAG	CCTTAGATAG	CATCCAATCT	AACATTTTT	31450
TGTGTGTTTG	AGAGAAGACT	ATCAAGAGCT	AGGAATAACT	TTTTAAAGGT	31500
AAGCCATTG	CACTATAGTG	TGGATTTTG	TTAAAAGGGG	ATAATTGAA	31550
ATTTTATGAC	TCATTATACA	AGACAAAATA	AGTTGGATT	TCAATGTTT	31600
TACAAAGTAA	ATCAAAGTTA	TAATTGCCAA	CAGTACCGAA	AGCTTCAAAA	31650
CATTTTTAT	GTATGAAAT	TGTAATT	TTAACCTTAA	AATGAGCCAG	31700
TACCATGTTG	TTGCTTAAA	ATCTCATGCT	AAGAATTCTAC	TATGTTGTTA	31750
ATAATCTTCA	AGATATT	GAATAAACT	TTATTCTCAA	TCCTCTCTCC	31800
AACCTPATCT	GGTGTCAAAT	CAGGAATGTT	TTCTTCCAA	AAAGCCTCGT	31850
GGAAAGATCTG	TATGTCTAA	TATATGTCAG	GGATAATACA	GATGTAGCCC	31900
TGCGAACGCT	GACCTTGATT	TTTATAGTCT	AAAATGTCAT	TTGCAGATAT	31950
CTATTTCTA	AGAATAATT	CTAAAAGAAT	TATTGTAATG	TTGAGGAAA	32000
GCTAAGAAAT	TTTGCAAAGA	GCCTAGTGA	AAATATAAGC	TAGGCTTTTG	32050
TGGTTTGTGG	ATAGACTTCC	CAACAAAATT	GCTTTTTATC	TATAGTGATC	32100
CAACCTTGTG	GAACATATT	GTCATCTTTT	TTTAGAAAAT	TCTTAGAAAA	32150
GTGATCTTGC	AAAATGGAA	TTTATCTTC	CCCAAGTATA	TTCTGTCATG	32200
TATAGAGTTA	AACAAAGCAT	AGTAATTCA	CCAGACAAAC	ATTCAAAATC	32250
TACTCCTGAC	CTTTTTATCT	CATCCAATT	TTCCCAGGGC	CCAGACATAA	32300
ACCTTTGCCT	TACGAACCT	TTGTATATGC	ACTAAATATG	CTTCTCCTTC	32350
AAGGTTCTCA	GTCACTAGA	AAAATGTCA	AGAGTAATATG	GTACCCCTTCT	32400

CACTTGTAGA	TCCAAGAGAA	TTAGACTTAA	ACTCACTCTA	CATGTCTGTG	32450
ACTTTATTTT	ATTTGCATGA	CAGTCCTGTG	AGGTGGCAAG	GCAGGTATCT	32500
TGGATCCATT	TTTAGATAA	GGAAGTCAA	ATTGAGAAGA	GGTTGCATGA	32550
TTTACAGGAA	GCCATACTGT	AGTCCTATGT	TACTCTTAA	AATCCCATTC	32600
AAATCCTGCT	TCTGAGGCCT	GCATACTTTC	TACCCCTACCA	GTCATTGACC	32650
CATGCTTATG	TCTCCTTGA	AAACATTGAT	TCCACTCTTG	TCTCCAGTGA	32700
AAAAGTGGAA	TTTAAGCAGA	GAAACAAAAG	CCATTGTCT	TGTTAAGTCT	32750
ACTTTCCCTC	TACTTTCAAG	AAGGAAAGTT	GGGGTATGTG	TTGAAATGGTG	32800
ATTTATTTAT	TTATTTATTA	TTTAAAGAT	TGATACAAGG	TCTTACTGTG	32850
TTGTGCAGGG	TGGTCTCAA	CTCCTGGCT	CAACTGATCA	TCCCACCTCA	32900
GCCCTCCAGT	GTTGGGATTA	CAGCATGAAC	CATTGTGCC	ACCAACGATC	32950
CGCAGTTTT	TAAGAAAAC	TTTTACTATA	GAAAATTATA	ATCATATACA	33000
AAATAACAGAG	GAAAGTATAT	GAACCCACTT	TAGGAGACTA	GAATATGCCA	33050
CCCCAAAATA	TGCCACTTTC	GCATAAAGGAT	TATTTCGAGC	TAAGGCAAC	33100
TGGAAGAAA	CACATAGAA	AAAAGTCTC	TGTCCTCTC	CATTGCTTA	33150
AAAGCAGGAC	ATGAATCTTA	AAAGTCCCCC	TCCCTCCCTT	TCTACCCAGGA	33200
AAAACAAGAG	TTAATCACTG	AAGATAACTT	CAGACCCCTTA	TCAGTGTAGA	33250
GATGGCACTA	GAAGAACTTA	TATTACATAC	TCATTATTT	TCCTTCCCAC	33300
AACTTGCCAC	CCCAGAGACT	AAAATCCTT	TTCCCTGTG	ATGTCCTTGT	33350
TCCAAAATT	TGCTCTATA	GCTGGAGTT	TAAGCCACCT	CTTGTAGAAT	33400
TACTTGTCC	CTGGTATTTT	CTGTTAACAT	ACATGTATTA	ATATACATGT	33450
TAACAAGCTT	CTGTTGTTT	TTCTCCTGTT	TTCTGTCTTG	TTACAGAGGT	33500
CCATCCCAAC	TAAGAACTAA	AGAGTAGGAG	GAAAATATAA	TTTCCCTCTG	33550
CATACTTGA	TCTTGTAA	TCCGTAAACC	TTCCCACCTT	TCACCTCCTA	33600
CCTATTAGAT	TACTTGTAA	CAAAATTTCAG	ATATATTACT	TTATCTATAA	33650
ATATTTCACT	ATGTGCTAGG	TGTGGTGGCT	CACACCTGTA	ATCCCAACAC	33700
TTTGGGAAGC	TGAGGCAGGA	GGATCACTTG	AGCCCAGGAG	TTCAAGACCA	33750
GCTACGGCAA	AAAAACTATCA	AAAACTATAC	TGGGCATGGT	GGCACATGCC	33800
TGTGGTCCCA	GCTACATGAG	AGGCTGAGGC	AGGAGGATCG	CTTTAGCCCA	33850
GGAGGTTGAG	GCTGCAGTAA	GCTGCATCA	CACCACTGCA	CTCCAGCTG	33900
GGTGACAGAG	TAAGACCATG	TCTCAAAAAA	ATACATATTT	TAGTATGTAT	33950
CCTTTTTGTA	AAAACACAAAT	ACTTTTATCA	TACTTTAAT	AATAACAATA	34000
ATTCCCTAGT	ATCACAAAT	ATTTGTCA	TGTCTCACAT	TTTCCCTTATT	34050
GTCTAAAATA	TTGTTGATAG	TTATTCAAA	CAGAATCAA	ACAAGTCCA	34100
TATATTACAT	TTGTTGACA	AGTCTTTAA	GTGTTGTCAT	CTTAAAGTTC	34150
TTCCCTCCCTC	TCTTCATCT	CTTGTAAATTT	ATTAATGTGA	AAAAACAGGT	34200
AATTGTTCT	ATAGTATTTC	CTACATTATA	GAGTTGCTA	CATTATTCTC	34250
CTATGATATC	ATTAGCATG	TTCCCTCTGTC	CCCTGTGTTT	CCTGTAAACT	34300
GGTAGTTATA	CCTAGAAGCT	TGAGTTTATT	CAGGTTTTA	ATTGTTATT	34350
TTTGTCAAGA	ATTTTCTTATT	ATCTGCTTCT	GGAAGCACAG	AATGTCGGT	34400
TGTGCTGGT	TTTGATCTTG	ACAGCTACTG	ATGACCATGG	CCTAATCCAT	34450
TACTTTATTG	GGGTGGGGGG	AATAAGGTTT	TAAAATTAAT	TTTTTTAAA	34500
GATTTTTTA	ACTGTTATTT	TGAGACAGTG	TCTCATTCG	TTTCCCAGGC	34550
TGGAGTGCAG	TGGCACAATC	ACGGCTCACT	GCAGCCTTGA	CCTCCTGGGA	34600
TCAGGTGATC	TTCTCACCTC	AGCCTCTGG	GTACCTGG	CTACAGGTGC	34650
ACACCAACAC	ACCTGGCTAA	TTTTTGTTAT	TTTGTGAC	GAAGGGTTT	34700
CATCATGTTT	CCCAGACTGG	TCTTGAACTC	CTGGCTTCAA	GTGATCTACC	34750
CACTTCAGCT	CCCCAAATC	CTGGGATTAC	ACTTTGGCCA	CCGTGCCTGG	34800
CCTAAATGAA	ATTATTGTC	TCTAACACAGA	CAGAAGTTTT	ACTTTAAAAAA	34850
TTTGTCTTTG	TGTGTACATG	TGTTGTGTA	TGTGTG	TCTAAAAGT	34900
TGGCTTGTAG	CTTGTCTTG	AATTCTGGA	TGAACAAATAA	CCAAGAATAC	34950
TTAAACTCTG	ATCATTCTG	ACAGATATCC	CCTACAGGCT	ATGGCCTTTT	35000
GAATTGTC	CTCCAGTGAT	AAAAACGAGC	AAGCACGATA	CTGCTCTCAG	35050
ATTCATGGTG	TCACATGTG	AGGTGAAAAA	AAAAAAAAG	ATGAATCCCA	35100
TTAAATGCC	CCCAGGATAA	CACTGATACT	CTTGTGAGGA	TAACTATTG	35150
CTTGCCTACTG	TTTCATTA	ATAAGGACAT	AAGTAAAGAT	CTATTTTGT	35200
CTCTTCTCC	CCAACCCAC	CAACTAGGAT	TATTGGCTAT	CTCTCTGTG	35250
CAAGAAATTG	GTGGGACCA	AGGTGTTAAT	GGCAAGCGTG	CAAGGTTCAA	35300
AGAGAAGGAA	GCTTCAGTA	TACCTTCATT	GCACAAACAC	TGACAAGTAA	35350
GTATGAAACA	CACCCCTTAC	CAATCATCAA	CTTTAGTGC	GTAAGCCTGT	35400
AACTTACTC	AAACACCCG	TTGCATGTG	CTATACATTG	CATAAGTATA	35450
GGCAGTGTCA	ATTAGTAAA	GTGTTATACA	ACGATTTTAT	TTTATTTTAT	35500
TTTTAGAAGA	AAAATGCTAC	TTTGTGTTT	GTTGTTTTT	GAGACGGGGC	35550
CTCGCTCGTC	ACCCAGGCTG	GAGTGCAGTG	GTGCAATCTC	AGCTCACTGC	35600
AACTCCGCC	TCCCCGGGTT	AAGTGAATTCT	TGAAGAGGAG	AAACATAATA	35650
ACAACAAAT	TATTTTCAAA	AGTTGTGACC	GCAGTTCTG	GAGTTGAGAA	35700
GACATCGAGA	TTTTGTAGC	CTCATACTCT	TGCTTAAAGT	AGCAAAAAAT	35750
GTTCCCTAAAT	CTCAGGAATA	TTCTCATGAT	AGGTTTCAA	CTATCATTC	35800
TGATAAGATG	ATGCTGAAAT	ACTAATTCTA	GCCAAAAAAG	ACCAGCTACC	35850
ATTCCGATT	TTGGGGGACT	GGGAACCTG	GATAGTGAGG	ACCCAGTAG	35900
GAAGTAGCGA	GGGAATGGT	TTGAATGGAT	AAATTCAAA	AAAATGTCAG	35950
TAGATTTAAT	TTTCTTATAC	ATTCAGTCT	TTTATATAAGG	CTAGGAAAAG	36000
CCCCGTGTTT	TATGGTTTAT	AATTGAATT	CACATGAACC	CACAAATT	36050
GCCTTTTAC	TTCTATGTC	TGAAATGGA	TAGTCTGGCT	GGCCTCTAA	36100
CAACCCAGCT	GGCAGAGCTG	TGAGGATCTC	AGTGTGCT	AGCCAGACA	36150

TTGGTAGCAT	GAACGGCAAC	ATTTTAATT	GTGTTTTCAA	AATAGGAGCA	36200
CACTAGCGGT	CTAAAACGAT	CATAAAAGAA	GGATACTAAG	AGGGCCCACT	36250
GTCATTATGG	ATCCTAATAC	TTAGGATGCA	TTATGGATTG	TCATTATGGA	36300
TACTAATACT	TAGGATCACA	TTTGTAAATTG	AGTTTTAAAT	TGCTTAAATT	36350
AGATACATAT	TTCTTATTAAAG	TTAACCTCTT	TGCTTTAGT	CCAAGGTATA	36400
AAGAAAGGAGA	TTTAACCTCG	TATGCCATAA	ACCTCCATAA	TGTCCACCAAG	36450
TACTTGCAGGT	TACCCCTATCC	TTTTCTAAC	AAGCAAGTGG	ATAAATACCT	36500
TCTAAGACCT	TTGGGACCTC	ATGGATTACT	TTCCAAGTAA	GTAAATTTC	36550
TTGTTCATTC	CAAACCTTCA	ATAAAATTAT	TGGTGTGTTAT	CAGAATAGAG	36600
AGTTGGACAC	GGGAGCAAA	GACAAAGTC	ACTATATCAA	GTCTAATAAA	36650
TTCTTAATAT	TCAGGAAATT	TATGTATGAA	TACTTAACTAA	TATGAGTATA	36700
ACTCATCCTA	AGAGTCTAAA	GCAAAAGGAT	GTGACACAAA	ACTAGCAGTT	36750
ATCTTAGAGA	ATAAGTTTGC	ATTTCAAAAT	AACTTGACAT	ATCAAGATCC	36800
ACTCAACGCA	TTTAAATTAT	TTACTCTAA	AAGACATAAT	TCTTGGTAAC	36850
ACATTCACTA	AAGCAAATA	TACCTTTATA	TAATTGCTAT	CAAGGTATG	36900
TGGGTTGGTA	AAAAATATCA	TACCATGTGA	GATCACTGTG	ATTCTTTAC	36950
AGCATTAAATT	TTTATTGGTT	AGAGTAAGAA	AAAGAATAGC	TAGAGTATAT	37000
TTCTTAAAGT	GATTCTCAT	CACTTTGGTT	TCAAAACCCA	ATTATTGACT	37050
ACATCTTATA	AAAGCCTGTA	TTCAATGGAG	TGCCCCAAA	TGACTATGAG	37100
TCTTAAAGAG	TAGGCATAT	AAATATTATA	AGGTTCTGT	TCATGTATG	37150
TTGGAAGGAG	TTCTTTCTC	ATGACTATT	TCATATTGGA	GCATAAAAAG	37200
AGTTTACAGG	CTTGGCGAG	TGGCTCATGC	CTGTAATCCC	AATACTTTGG	37250
GAAGCTGAAG	CAGGCGAGTC	ACTTCAGCCC	AGGAGTTTG	GACCCAGCTG	37300
GGCAATATGG	CAAACCTCTC	TCTACAAAAT	ATACCAAAAT	TAGCCAGGCG	37350
TGGTGGTGCA	TGCGCTGAGT	CCCAGCTACT	TGGGAAGCTG	AGGTGGGAGG	37400
ATTGCTTGAG	CCCAGGGGGG	TCATGGCTGC	AGTGAAGCTGT	GATGGTGCCT	37450
CTGTCAACCA	GCCTGGGTGA	CAGACTGAGA	CCCTGTCTCA	AAAAAATAAA	37500
TAATAAAAAA	TTAAGAGTTT	ACAAAATTCT	CACCATCTCC	TCCCACATCTT	37550
GCAAATGCCA	CATAACTGAT	GTGTTCCAGG	ACTATTAGCC	TCGGAACCTG	37600
AGGCAGTACA	GTAAGCACGC	TTTCTCCAAA	GTCCTGTCCC	CCACAGACAA	37650
ACATTATTTA	CACTGGGTAC	TGCTCTTTTA	TTTTTCCCC	TCTATGCTTT	37700
ATTTTACTAT	AACATAATACT	ATATAACATG	TAATAGGAA	AAGGCAGGGT	37750
CGGGGGAGAG	ATCCAGAAGT	CTTCCCAAGA	GCCTTTCCAA	CATAGCCTCT	37800
GTAGACATT	TTTCTTTCTT	CTTTTTTTT	TTTTTTTTT	TTCTGAGACA	37850
GAGTCTCACT	CTGTTGTC	GGCTAGAGTG	CAGTGGCGTG	ATCTAGGCTC	37900
ACTGCAACCT	CCGCCCTCTG	GGTCAAGCA	ATTCTCCCAC	CTCAGCCTCC	37950
CTAGTAGCTG	GGATTAGAGG	CATGCATCAC	CACGCCCTGC	TAATTTTG	38000
ATTTTAGTA	GAGATGAGGT	TTCACCATGT	GGGCCAGGT	GGTCTTGAC	38050
TCCTGACCTC	AAAGTATCCA	CCTGCCTTAG	CCTCCCAAAG	TGCTAGGATT	38100
ACACGAGTGA	GCCACCGTGT	CTGGCCCTCTA	TTACATCTG	ATCACACATT	38150
TCATGTTTA	TAATTGAAA	ACTGGTAAA	TTATAGACAA	TGTTTGTTC	38200
CCCTAAATT	TCTTGTATGA	GTATATATTA	CTTACACTCT	TCTGTCTTTA	38250
AAATTTTGCA	AAATAGTATC	CTAGATAAGT	TTATGAGTGC	ACAGTCTGTA	38300
CGCTTACTCA	TATTAATGAC	CTCGGAGAGT	AAACAAACAG	TCACCTTAA	38350
AAATTATTAC	TATCATTACT	ATTATTTTG	AGGCGGGGT	CTCATCTGT	38400
CTTCCCAGGCT	GGAGAGTAGT	GGTGCCTGCA	CAGCTACTG	CAGCCACCGC	38450
TACCTGGGCT	CAAGTATGCC	TTCCCTCTCA	GCCTCTGAG	TAGCTGAGAC	38500
CACAGGCTTA	TGCTTACCA	CTGGCTAA	TTTTTAACTT	TTTGTAGAGA	38550
CGATGTCTCA	TTATGTTGCC	CAGGCTGTC	TCAAACCT	AAGCTCAAGT	38600
GATCTTCTC	AGCCTCCCAA	AGTGTGGGA	TTACAGGCAT	AAAAAACTGC	38650
ACCCAGCCCT	AAAATTATT	AGGGTCTGC	ATAGTAAGAC	TTAATAAAAT	38700
ATTAAATGAA	ACATCTGGTT	TTTTTAAAAA	AAAAATAGAG	ACAAGGCTC	38750
ACTATATTGC	CCAAGCTGGT	CTCGAACTCC	TGGACTCACG	CAATCCCTGCT	38800
GCCTTAGCCG	CCCCAAGTGC	TGGGATTACA	GGCATGACCC	ACCTCATCTG	38850
GGCTGAGTGA	ACATATT	AAACATAAAGG	CCGTATT	TATTGATCTC	38900
ATACATTG	CCCAGCATCC	CCATTTCGGC	CGAATCTGT	GCTTGTAAAT	38950
TCCTTCCAGC	TTCATTTCAT	CTGAAATTG	ACAAACATCT	TCTATTCTT	39000
TGTGCTCATG	TTATTGACTT	CAGAAATAA	AAATAAACAC	TATAACCAA	39050
TTAAACCCCCA	CCCTCATTC	CCAGCCTGT	GTGAAAATAA	TCAGCATACA	39100
TTAAGCTTAC	CCTTGATATA	TGTGTAGCAT	CTTTAGATA	AAATACAGC	39150
TGATTAAGCA	ATATAGCC	ATGGTATAAT	ATCTTGCCC	TGTACCTCAT	39200
CTTATCTCCA	GCAGGATTA	TTCACAGTG	TCAGATTAC	CTTTAAACTT	39250
TGTAGCAAAA	TATCTCTCC	AAAAGCAT	CTAAAATT	TGTGTGTACT	39300
CTTGAAGCTT	TCTTAATTTC	ATGCAGAAC	GGCTCTTAC	ACTGTTAGCT	39350
GGAGATATT	TCAAGACCTA	TTTTGTTG	TGGTTCTG	ATGATGGTC	39400
TGGCATT	CCCTTCACTC	CATCTAAAAA	TTGAGGTGAT	ACAGGCTTTT	39450
AAACAAAACC	AACTCATATA	GACTGAGTAC	AACTGCAATG	CAGGCATGCT	39500
AACTCTGCT	ACAATCATGG	GCCTGCTATT	GATATGTCTT	AAAGTTACAGA	39550
ACACAGGGCT	GAGCGTCTCA	TTAGGTCAA	ATGAAACCA	GTTTTCTG	39600
TCACTGATGC	TTAATGAGGA	CAGGGTGTGA	GAGATTCTT	TAAGGAAAAC	39650
AAATATATAA	TAATGCTACA	TGGAAAATA	TCTAACATTA	GAGAATTAAG	39700
TAATAAAAT	ATATAC	CACCATGAA	TCTTGTGAG	ACATTTAAAT	39750
TATGTAGTGG	ATGGATGTTT	AAAGGTGTGA	AAAAAGTTA	GGATGTGCTG	39800
GGTGGGGGG	AAGAATCAAG	TTTAAGAAA	ATACAGTATA	CCCATACTTA	39850
AGTAAAAAAA	AAAAAAAGG	TATGTACAGT	CATGTGTTGC	TTAATGATGG	39900

GGATACATTC	CGAGAAAATGT	GTCGATAGGT	GATTTCATCC	TTGTGTGAAC	39950
ATCATAGAGT	GAACTTACAC	AAACCTAGAT	GGCTAGCCT	ACTATGTATC	40000
TAGGCTATAT	GACTAGCCTG	TTGCTCTTAG	GCTACAAACC	TGTAAGGCAT	40050
GTTACTGTAG	CGAATATACA	AATACTTAAC	ACAATGGCA	GCTATCATTG	40100
TGTTAAAGTAG	TTGTGTATCT	AAACATATCT	AAAACATAGA	AAACTAATGT	40150
GTTGTGCTAC	AATGTTACAA	TGACTATGAC	ATTGCTAGGC	AATAGGAATT	40200
ATAAATTTAT	CCTTTTATGG	AACCACACTT	ATATATGCGG	TCCATGGTGG	40250
ACCAAAACAT	CCTTATGTGG	CATATGACTG	TATACATGTA	CACAAAAAAAT	40300
AGATGAAAGA	ATGAATATAC	ATCAAAATAT	TTAAAATGGT	TATAATGACT	40350
TAGGITACCTT	TTATTTATCT	TAGTAATATA	AATGATGATA	GATAATACCTT	40400
TTATAGTGT	TACTATATAA	AAGACACTGT	TATAAGTGT	CTACATACTT	40450
TACATGTATT	ACCTAAATGA	TATAAATATA	ACTCTGACAG	TAACAAATCT	40500
TATACGTCT	CTTTTCTTTT	TTTTTTTTT	CTTTTTTTAG	ACAGAATCTT	40550
GCTCTTACAG	GCTGGACTGC	AGGGTGCAT	CTCGGCTCAC	TGCAACCTCC	40600
GCTCTCCAGG	TTCACAAACGAT	TCTCATGCT	CAGCCTCTC	AGTAGCTGGG	40650
ACTACAGGCA	CACACCAACCA	TGCCCCGCTA	ATTTTTGTAT	TTTGGGTAG	40700
AGATGGAGTT	TTGCCATGTT	GGCCAGGCTG	ATCTTGAAC	CCTGGCCTCA	40750
AGTGATCTGC	CTGCCCTCAGC	CTCCCCAAGT	GCTGGGATTA	CAGGTGTGAA	40800
CCACTGTGCT	CGGGCTAATC	TTACAAGTTT	TCAATATTTA	AAGAGTGCCTA	40850
ACTTTGTTGA	CAATATAAA	CATATTTGAG	AAAAAGAGAT	ATAAGCATCT	40900
TATTTAGAAT	TATGAAAATA	TCAATAGACC	TACAGCCGAC	TAAAGCTTTT	40950
CTTCATAAGC	TCTTGCTAT	ATTGATTGCG	TCCTGTGAAT	ATGCATTAAT	41000
TTGATTTAA	TAATAAGTAT	GTATAAGAAA	TAACACTTTT	CCTTAATTTT	41050
TAAGAACGTT	CAACAGTTT	TAATTTGAA	TCCAATAGT	AAATACATAG	41100
AAAAATATATAA	ATTTTCTGTA	TTTAGGCAA	ATTGTTTTG	TTTACCCACA	41150
GCATTCTACC	AAAATTCTT	ATAAACAGTA	AGAAAATGAA	TGCATACCTC	41200
CTGCAGGGAG	AGGGGAGTTA	GGCAGTTTAT	GGGCATAGTT	ACAAGTGAGA	41250
AATTTCATTG	GCTACCATT	ACGCTAAATT	CATAAAAAC	GCATTCAATT	41300
CTATATATCT	ATTTTCTTTA	CATAAAAAAAG	GTTTCAATTA	TTGCCATTAA	41350
AATAAAATAG	CCACCATTCC	AGAAGTTGT	TCATGTTAT	CTTTTTTATA	41400
CCACCATCAT	ATTGCCTATT	ATATAGATTG	TGTGTGTTCC	ATTTTCTGTA	41450
ATGGGCCAGA	CAGTAAGTAT	TTCTGGCTTT	GGAGTCCATA	TGGCTCTAT	41500
CATAACTACT	CATCTCTGCC	ATTGAGCTT	AAAGATTATC	TAGGTCAAAAT	41550
GCTTAAGTGA	TATAGTGTG	AAATACAAT	TATATAATAT	AGGCTGCCAC	41600
AAAAAAAAT	TTATTTGGTC	AAAAAAAGAT	TTCATGACTT	TTGAGCAGC	41650
ATGGGTGGGG	CATGCACCAC	TTGGTTAACT	CGGTGTATCT	TTCTCTTTG	41700
CAGATCTGTC	CAACTCAATG	GCTAACTCT	AAAGATGGT	GATGATCAAA	41750
CCTTGGCACC	TTTAATGGAA	AAACCTCTC	GGCCAGGAG	TTCACTGGGC	41800
TTGCCAGCTT	TCTCATATAG	TTTTTTGTG	ATAAGAAATG	CCAAAGTTGC	41850
TGCTTGCATC	TGAAAATAAA	ATATACTAGT	CCTGACACTG	AATTTTCAA	41900
GTATACTAAG	AGTAAAGCAA	CTCAAGTTA	AGGAAGAGAA	GCAGATACCT	41950
TGCAAAGCAA	CTAGTGGGTG	CTTGAGAGAC	ACTGGGACAC	TGTCAGTGC	42000
AGATTTAGCA	CAGTATTTTG	ATCTCGCTAG	GTAGAACACT	GCTAATAATA	42050
ATAGCTAATA	ATACCTTGT	CCAAATACTG	CTTAGCATTT	TGCACTGTTT	42100
ACTTTATCT	AAAGTTTTGT	TTGTTTTAT	TATTTATTA	TTTATTATTT	42150
TTGAGACAGA	ATCTCTCT	GTCAACCCAGG	CTGGAGTGCC	ATGGTGCAG	42200
CTTGGCTCAC	TGCAACTTTA	AGCAATTCTC	CTGCTCTAGC	TTCTTGAGTA	42250
GCTGGGATTA	TAGGCGTGTG	CCACCAAGCC	CAGCTACTTT	CTATATTTT	42300
TGTAGAGATG	GAGTTTCGCC	ATATTGCCA	AGCTGGCTC	GAACCTCTGT	42350
CCTCTGAACTC	CTGTCCTCAA	GTGATCCACC	CGCCTCAGCC	TCTCAAAGTG	42400
CTGGGATTAC	AGGTGTGAGC	CACCAACCC	AGCAGTTTT	TATTTTGAG	42450
ACAGGGTATC	ATTCTGTTGC	CCAGGCTTGA	GTGCAGTGGT	GCAATCATAG	42500
ATCACTGCAG	CCTTTTAACT	CCTGGGCTCA	AGTCATCCTC	CTGCTTAGCC	42550
TCCCAAGTCT	CTAGGACCAC	AGACACATGC	CATCACACTT	GGCTATTTT	42600
AAAAATTTT	TTGTAGAGAT	GGGGCTCGC	TATGTTACCC	AAACTGGTCC	42650
TGAACCTCTG	GACTCAATTG	ATCCCTCCAC	CTTGGCCTTC	CAGGTGCTGG	42700
GATTTTTTG	GGAGTACAGC	ATGGTACAGC	AGGAGATCAT	TTGATGTTAC	42750
CTCTGTGCG	TGTTGCTAGT	CAGCGAAAGA	CTATAATACC	TGTGGGGACA	42800
GCGGATTAGCC	ACCACACCCA	GTCTTTATT	AAAGTTATTA	AAAATGGCTG	42850
GGCGCAGTGG	CTCACACCTG	TAATCCTAGC	ACTTTGGGAG	GCCGAGGCAG	42900
ATGGATCACC	TGACGTGAGG	AATTGAGAC	CAGCCTGGCC	AACATGGTGA	42950
AACCCCACT	CTACTAAAAA	ATACAAAAAT	TAGCTGGGT	TGGTCCCTGTA	43000
GTCCCGAGCT	CTTGGGAGGC	TTGGGAGGA	GAATTACTTG	AACCCAGGAG	43050
GCAGAGGTG	CAGTGAGCCG	AGATTGTC	ACTGCACCTC	AGCCTGGGTG	43100
ACAGAGAGAG	ATTCCATCTC	AAAAAAACAA	GTATTAAAAA	ATGTATATGA	43150
ATGCTCTAA	TATGGTCAGG	AACCAAGGAA	GGCAAGGATA	TATTATGAGT	43200
TTTAAGAAGG	TGCTTAGCT	TATTTATAC	TTTCAAAATG	TATTAAGAAGA	43250
TTTTAGAATT	CTTTCCTTCA	TGTGCCATCT	CTACAGGCAC	CCATCAGAAA	43300
AAGCATACTG	CGGTACCGT	GAAACTGGTT	GTAAAAGAGA	AACTATCTAT	43350
TTGCACCTTA	AAAGACAGCT	AGATTTTGCT	GATTTTCTTC	TTTCGGTTTT	43400
CTTGTGCTAGC	AATAATATGT	GAGAGGACAG	ATTGTTAGAT	ATGATAGTAT	43450
AAAAATGGT	TAATGACAAT	TCAGAGGCAG	GGAGATTCTG	TAAACTTAAA	43500
ATTACTATAA	ATGAAATTGA	TTTGTCAAGA	GGATAAATT	TAGAAAACAC	43550
CCAATACCTT	ATAACTGTCT	GTTAATGCTT	GCTTTTCTC	TACCTTTCTT	43600
CCTTGTCTCA	GTTGGGAAGC	TTTGGCTGC	AAGTAACAGA	AACTCCTAAT	43650

TCAAATGGCT	TAAGCAATAA	GGAAATGTAT	ATTCCCACAT	AACTAGACGT	43700
TCAAACAGGC	CAGGCTCCAG	CACTTCAGTA	CGTCACCAAGG	GATCTGGTT	43750
CTTCCCAGCT	CTCTGCTCTG	CCATCTTAG	CGCTGGCTTC	ATTCTCAGAC	43800
TCTGGTAGCA	TGATGGCTGT	AGCTGTTCA	TGGGGCCCTT	CAAACCTCAT	43850
AGCAACCAGA	GGAAAGAAAAT	GAGCCATT	TTGAGCTCTC	TTCATAGACT	43900
TGAATAACTC	TTTTTCAGAG	CTTCTCACAG	CAAACCTCTC	CTCATGTCTC	43950
CTCATGTCCT	ATTGTTCAAGA	AATGGGTAAAT	GTGGCCATT	CACCAAGTCAC	44000
TGCCAACAAAC	AACGAGGTTC	CTATAATTGT	CTCTGAGTAA	CCCTTGGAA	44050
TGGAGAGGGT	GTTGGTCAGT	CTACAACTG	AAACACTGCG	TTCTGCGCTT	44100
TTTACCACTG	AAAAAAATGTA	ATTATTTCC	CCTCTTAAGG	ATTAATATT	44150
TCACAAATGTA	TGCGCTGTTAT	GGATATAGTA	TCTTTAAAT	TTTTTATTTT	44200
AATAGCTTCA	GGGGTACACA	CTTTTGCTT	ACAGGGTGA	ATTGTGTAGT	44250
GGTGAAGACT	CGGGCTTTAA	TGTACTTGTC	ACCTGAGTGA	TGTACATTGT	44300
ACCCAATAGG	TAATTTTCA	TCCATTACCC	TCCTTCCGCC	CTCTTCCCTT	44350
CTGAGTCTCC	AACATCCCTT	ATACCACTGT	GTATGTTCTT	GTGTACCTAC	44400
AGCTAAGCTT	CCACTTATAAT	GTGAGAACAT	GCAGTATT	GTGTTCCATT	44450
CCTGAGTCTAC	TTCCTTCTAGG	TAACACGCC	CCAGTCCCGT	CCAAGTTGCT	44500
GCAAAATACA	TTATTCCTCT	TTATGCTGA	GTAATAGTCC	ATGGTACATA	44550
TATACCACAT	TTTCTTCTAC	CACTTATCAG	TTGATGGACA	CTTAGGTTAA	44600
TTCATTCAA	TTTCATTCAA	TTTAAGTATA	TTTGTAGGA	GCTAAAGCTG	44650
AAAATTAAT	TTTAGATCTT	TCACATCT	AAATTTTAT	ATGTAAGTGG	44700
TTTTTATATT	TTCACATTTG	AAATAAAGTA	ATTTTTATAA	CCTTGATATT	44750
GTATGACTAT	TCTTTTAGTA	ATGTAAGGCC	TACAGACTCC	TACATTGGA	44800
ACCACTAGTG	TGTTGTTCA	CCCCTGTTA	TACTATCAGG	ATCCTCGA	44848

(2) INFORMATION FOR SEQ ID NO:43:

(i) SEQUENCE CHARACTERISTICS:

(A)	LENGTH:	2396
(B)	TYPE:	nucleic acid
(C)	STRANDEDNESS:	double
(D)	TOPOLOGY:	linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:43

TTCTCTAGTTG	CTTTTAGCCA	ATGTCGGATC	AGGTTTTCA	AGCGACAAAG	50
AGATACTGAG	ATCCTGGGCA	GAGGACATCC	TAGCTCGGTC	AGATTTGGC	100
AGGCTCAAGT	GACCACTGTC	TTAAGGCAGA	AGGGAGTCGG	GGTAGGGTCT	150
GGCTGAACCC	TCAACCGGGG	CTTTTAACTC	AGGGTCTAGT	CCTGGCGCCA	200
AATGGATGGG	ACCTAGAAAA	GGTGACAGAG	TGCGCAGGAC	ACCAAGAACG	250
TGGTCCCACC	CCTGCGCGGC	TCCCGGGCGC	TCCCTCCCA	GGCCTCCGAG	300
GATCTTGGAT	TCTGGCCACC	TCCGCACCT	TTGGATGGGT	GTGGATGATT	350
TCAAAAGTGG	ACGTGACCGC	GGCGGAGGGG	AAAGCAGCA	CGGAAATGAA	400
AGAGAGCGAG	GAGGGGAGGG	CGGGGAGGG	AGGGCGCTAG	GGAGGGACTC	450
CCGGGAGGGG	TGGGGAGGGAT	GGAGCGCTGT	GGGAGGGTAC	TGAGTCCTGG	500
CGCCAGAGGC	GAAGCAGGAC	CGGTTGCAGG	GGGCTTGAGC	CAGCGCGCCG	550
GCTGCCCCAG	CTCTCCCGC	AGCGGGCGGT	CCAGCCAGGT	GGGATGCTGA	600
GGCTGCTGCT	GCTGTGGCTC	TGGGGGCC	TCGGTGCCT	GGCCCAGGGC	650
GCCCCCGCGG	GGACCGCGCC	GACCGACGAC	GTGGTAGACT	TGGAGTTTA	700
CACCAAGCGG	CCGCTCCGAA	GCCTGAGTCC	CTCGTCCCTG	TCCATCACCA	750
TCGACGCCAG	CCTGGCCACC	GACCCCGCCT	TCCCTCACCTT	CCTGGGCTCT	800
CCAAGGCTCC	GTCGCTCTGGC	TAGAGGCTTA	TCTCCTGCAT	ACTTGAGATT	850
TGGCGGCACA	AAGACTGACT	TCCTTATTTT	TGATCCGGAC	AAGGAACCGA	900
CTTCCGAAGA	AAGAAAGTTAC	TGGAAATCTC	AAGTCAACCA	TGATATTGCA	950
AGGTCTGAGC	CGGTCTCTGC	TGCGGTGTTG	AGGAACACTCC	AGGTGGAATG	1000
GCCCTTCCAG	GAGCTGTTGC	TGCTCCGAGA	GCAGTACCAA	AAGGAGTTCA	1050
AGAACAGCAC	CTACTCAAGA	AGCTCAGTGG	ACATGCTCTA	CAGTTTGCC	1100
AAAGTGCTCGG	GGTTAGACCT	GATCTTGGT	CTAAATGCGT	TACTACGAAC	1150
CCCAGACTTA	CGGTGGAACA	GCTCCAACGC	CCAGCTCTC	CTTGACTACT	1200
GCTCTTCCAA	GGGTTATAAC	ATCTCCTGGG	AACTGGCAA	TGAGGCCAAC	1250
AGTTTCTGGA	AGAAAGCTCA	CATTCTCATC	GATGGGGTGC	AGTTAGGAGA	1300
AGACTTTGTG	GAGTTGCATA	AACTTCTACA	AAGGTCAAGCT	TTCCAAAATG	1350
CAAAACTCTA	TGGTCCTGAC	ATCGGTCA	CTCGAGGGAA	GACAGTTAAA	1400
CTGCTGAGGA	GTTCCTGAA	GGCTGGCGGA	GAAGTGATCG	ACTCTCTTAC	1450
ATGGCCTCAC	TATTACTTGA	ATGGACGCAT	CGCTACCAAA	GAAGATTTTC	1500

TGAGCTCTGA	TGCGCTGGAC	ACTTTTATT	TCTCTGTGCA	AAAAATTCTG	1550
AAGGTCACTA	AAGAGATCAC	ACCTGGCAAG	AAGGTCTGGT	TGGGAGAGAC	1600
GAGCTCAGCT	TACGGTGGCG	GTGCACCCCT	GCTGTCCAAC	ACCTTTGCAG	1650
CTGGCTTTAT	GTGGCTGGAT	AAATTGGGCC	TGTCAGCCCA	GATGGGCATA	1700
GAAGTCGTGA	TGAGGCAGGT	GTTCTTCGGA	GCAGGCAACT	ACCACTTAGT	1750
GGATGAAAAC	TTTGAGCCTT	TACCTGATTA	CTGGCTCTCT	CTTCTGTTCA	1800
AGAAAACCTGGT	AGGTCCCAGG	GTGTTACTGT	CAAGAGTGAA	AGGCCAGAC	1850
AGGAGCAAAC	TCCGAGTGT	TCTCCACTGC	ACTAACGTCT	ATCACCCACG	1900
ATATCAGGAA	GGAGATCTAA	CTCTGTATGT	CCTGAACCTC	CATAATGTCA	1950
CCAAGCATT	GAAGGTACCG	CCTCCGTTGT	TCAGGAAACC	AGTGGATACG	2000
TACCTTCTGA	AGCCTTCGGG	GCCGGATGGA	TTACTTTCCA	AATCTGTCCA	2050
ACTGAACGGT	CAAATTCTGA	AGATGGTGGG	TGAGCAGACC	CTGCCAGCTT	2100
TGACAGAAAA	ACCTCTCCCC	GCAGGAAGTG	CACTAAGCCT	GCCTGCCTTT	2150
TCCTATGGTT	TTTTTGTCA	AAGAAATGCC	AAAATCGCTG	CTTGTATATG	2200
AAAATAAAAG	GCATACGGTA	CCCCTGAGAC	AAAAGCCGAG	GGGGGTGTTA	2250
TTCATAAAAAC	AAAACCTAG	TTTAGGAGGC	CACCTCCTG	CCGAGTTCCA	2300
GAGCTTCGGG	AGGGTGGGGT	ACACTTCAGT	ATTACATTCA	GTGTGGTGT	2350
CTCTCTAAGA	AGAATACTGC	AGGTGGTGAC	AGTTAATAGC	ACTGTG	2396

(2) INFORMATION FOR SEQ ID NO:44:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 535
- (B) TYPE: amino acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:44

Met	Leu	Arg	Leu	Leu	Leu	Trp	Leu	Trp	Gly	Pro	Leu	Gly	Ala	
5													15	
Leu	Ala	Gln	Gly	Ala	Pro	Ala	Gly	Thr	Ala	Pro	Thr	Asp	Asp	Val
	20												30	
Val	Asp	Leu	Glu	Phe	Tyr	Thr	Lys	Arg	Pro	Leu	Arg	Ser	Val	Ser
	35												45	
Pro	Ser	Phe	Leu	Ser	Ile	Thr	Ile	Asp	Ala	Ser	Leu	Ala	Thr	Asp
	50												60	
Pro	Arg	Phe	Leu	Thr	Phe	Leu	Gly	Ser	Pro	Arg	Leu	Arg	Ala	Leu
	65												75	
Ala	Arg	Gly	Leu	Ser	Pro	Ala	Tyr	Leu	Arg	Phe	Gly	Gly	Thr	Lys
	80												90	
Thr	Asp	Phe	Leu	Ile	Phe	Asp	Pro	Asp	Lys	Glu	Pro	Thr	Ser	Glu
	95												105	
Glu	Arg	Ser	Tyr	Trp	Lys	Ser	Gln	Val	Asn	His	Asp	Ile	Cys	Arg
	110												120	
Ser	Glu	Pro	Val	Ser	Ala	Ala	Val	Leu	Arg	Lys	Leu	Gln	Val	Glu
	125												135	
Trp	Pro	Phe	Gln	Glu	Leu	Leu	Leu	Leu	Arg	Glu	Gln	Tyr	Gln	Lys
	140												150	
Glu	Phe	Lys	Asn	Ser	Thr	Tyr	Ser	Arg	Ser	Ser	Val	Asp	Met	Leu
	155												165	
Tyr	Ser	Phe	Ala	Lys	Cys	Ser	Gly	Leu	Asp	Leu	Ile	Phe	Gly	Leu
	170												180	
Asn	Ala	Leu	Leu	Arg	Thr	Pro	Asp	Leu	Arg	Trp	Asn	Ser	Ser	Asn
	185												195	
Ala	Gln	Leu	Leu	Leu	Asp	Tyr	Cys	Ser	Ser	Lys	Gly	Tyr	Asn	Ile
	200												210	
Ser	Trp	Glu	Leu	Gly	Asn	Glu	Pro	Asn	Ser	Phe	Trp	Lys	Lys	Ala
	215												225	
His	Ile	Leu	Ile	Asp	Gly	Leu	Gln	Leu	Gly	Glu	Asp	Phe	Val	Glu

32

230	235	240
Leu His Lys Leu Leu Gln Arg Ser Ala Phe Gln Asn Ala Lys Leu		
245	250	255
Tyr Gly Pro Asp Ile Gly Gln Pro Arg Gly Lys Thr Val Lys Leu		
260	265	270
Leu Arg Ser Phe Leu Lys Ala Gly Gly Glu Val Ile Asp Ser Leu		
275	280	285
Thr Trp His His Tyr Tyr Leu Asn Gly Arg Ile Ala Thr Lys Glu		
290	295	300
Asp Phe Leu Ser Ser Asp Ala Leu Asp Thr Phe Ile Leu Ser Val		
305	310	315
Gln Lys Ile Leu Lys Val Thr Lys Glu Ile Thr Pro Gly Lys Lys		
320	325	330
Val Trp Leu Gly Glu Thr Ser Ser Ala Tyr Gly Gly Ala Pro		
335	340	345
Leu Leu Ser Asn Thr Phe Ala Ala Gly Phe Met Trp Leu Asp Lys		
350	355	360
Leu Gly Leu Ser Ala Gln Met Gly Ile Glu Val Val Met Arg Gln		
365	370	375
Val Phe Phe Gly Ala Gly Asn Tyr His Leu Val Asp Glu Asn Phe		
380	385	390
Glu Pro Leu Pro Asp Tyr Trp Leu Ser Leu Leu Phe Lys Lys Leu		
395	400	405
Val Gly Pro Arg Val Leu Leu Ser Arg Val Lys Gly Pro Asp Arg		
410	415	420
Ser Lys Leu Arg Val Tyr Leu His Cys Thr Asn Val Tyr His Pro		
425	430	435
Arg Tyr Gln Glu Gly Asp Leu Thr Leu Tyr Val Leu Asn Leu His		
440	445	450
Asn Val Thr Lys His Leu Lys Val Pro Pro Pro Leu Phe Arg Lys		
455	460	465
Pro Val Asp Thr Tyr Leu Leu Lys Pro Ser Gly Pro Asp Gly Leu		
470	475	480
Leu Ser Lys Ser Val Gln Leu Asn Gly Gln Ile Leu Lys Met Val		
485	490	495
Asp Glu Gln Thr Leu Pro Ala Leu Thr Glu Lys Pro Leu Pro Ala		
500	505	510
Gly Ser Ala Leu Ser Leu Pro Ala Phe Ser Tyr Gly Phe Phe Val		
515	520	525
Ile Arg Asn Ala Lys Ile Ala Ala Cys Ile		
530	535	

(2) INFORMATION FOR SEQ ID NO:45:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 2396
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: double
- (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:45

TGC TTT TAG CCA ATG TCG GAT CAG GTT TTT CAA GCG ACA AAG AGA	TT TCT AGT	8
TAC TGA GAT CCT GGG CAG AGG ACA TCC TAG CTC GGT CAG ATT TGG		53
GCA GGC TCA AGT GAC CAG TGT CTT AAG GCA GAA GGG AGT CGG GGT		98
AGG GTC TGG CTG AAC CCT CAA CCG GGG CTT TTA ACT CAG GGT CTA		143
GTC CTG GCG CCA AAT GGA TGG GAC CTA GAA AAG GTG ACA GAG TGC		188
GCA GGA CAC CAG GAA GCT CCC ACC CCT GCG CGG CTC CCG GGC		233
		278

33

GCT CCC TCC CCA GCC CTC CGA GGA TCT TGG ATT CTG GCC ACC TCC	323
GCA CCC TTT GGA TGG GTG TGG ATG ATT TCA AAA GTG GAC GTG ACC	368
GCG GCG GAG GGG AAA GCC AGC ACG GAA ATG AAA GAG AGC GAG GAG	413
GGG AGG GCG GGG AGG GGA GGG CGC TAG GGA GGG ACT CCC GGG AGG	458
GGT GGG AGG GAT GGA GCG CTG TGG GAG GGT ACT GAG TCC TGG CGC	503
CAG AGG CGA AGC AGG ACC GGT TGC AGG GGG CTT GAG CCA GCG CGC	548
CGG CTG CCC CAG CTC TCC CGG CAG CGG GCG GTC CAG CCA GGT GGG	593
ATG CTG AGG CTG CTG CTG TGG CTC TGG GGG CCG CTC GGT GCC	638
Met Leu Arg Leu Leu Trp Leu Trp Gly Pro Leu Gly Ala	
5 10 15	
CTG GCC CAG GGC GCC CCC GCG GGG ACC GCG CCG ACC GAC GAC GTG	683
Leu Ala Gln Gly Ala Pro Ala Gly Thr Ala Pro Thr Asp Asp Val	
20 25 30	
GTA GAC TTG GAG TTT TAC ACC AAG CGG CCG CTC CGA AGC GTG AGT	728
Val Asp Leu Glu Phe Tyr Thr Lys Arg Pro Leu Arg Ser Val Ser	
35 40 45	
CCC TCG TTC CTG TCC ATC ACC ATC GAC GCC AGC CTG GCC ACC GAC	773
Pro Ser Phe Leu Ser Ile Thr Ile Asp Ala Ser Leu Ala Thr Asp	
50 55 60	
CCG CGC TTC CTC ACC TTC CTG GGC TCT CCA AGG CTC CGT GCT CTG	818
Pro Arg Phe Leu Thr Phe Leu Gly Ser Pro Arg Leu Arg Ala Leu	
65 70 75	
GCT AGA GGC TTA TCT CCT GCA TAC TTG AGA TTT GGC GGC ACA AAG	863
Ala Arg Gly Leu Ser Pro Ala Tyr Leu Arg Phe Gly Gly Thr Lys	
80 85 90	
ACT GAC TTC CTT ATT TTT GAT CCG GAC AAG GAA CCG ACT TCC GAA	908
Thr Asp Phe Leu Ile Phe Asp Pro Asp Lys Glu Pro Thr Ser Glu	
95 100 105	
GAA AGA AGT TAC TGG AAA TCT CAA GTC AAC CAT GAT ATT TGC AGG	953
Glu Arg Ser Tyr Trp Lys Ser Gln Val Asn His Asp Ile Cys Arg	
110 115 120	
TCT GAG CCG GTC TCT GCT GCG GTG TTG AGG AAA CTC CAG GTG GAA	998
Ser Glu Pro Val Ser Ala Ala Val Leu Arg Lys Leu Gln Val Glu	
125 130 135	
TGG CCC TTC CAG GAG CTG TTG CTG CTC CGA GAG CAG TAC CAA AAG	1043
Trp Pro Phe Gln Glu Leu Leu Leu Arg Glu Gln Tyr Gln Lys	
140 145 150	
GAG TTC AAG AAC AGC ACC TAC TCA AGA AGC TCA GTG GAC ATG CTC	1088
Glu Phe Lys Asn Ser Thr Tyr Ser Arg Ser Ser Val Asp Met Leu	
155 160 165	
TAC AGT TTT GCC AAG TGC TCG GGG TTA GAC CTG ATC TTT GGT CTA	1133
Tyr Ser Phe Ala Lys Cys Ser Gly Leu Asp Leu Ile Phe Gly Leu	
170 175 180	
AAT GCG TTA CTA CGA ACC CCA GAC TTA CGG TGG AAC AGC TCC AAC	1178
Asn Ala Leu Leu Arg Thr Pro Asp Leu Arg Trp Asn Ser Ser Asn	

	34		
185	190	195	
GCC CAG CTT CTC CTT GAC TAC TGC TCT TCC AAG GGT TAT AAC ATc Ala Gln Leu Leu Leu Asp Tyr Cys Ser Ser Lys Gly Tyr Asn Ile 200	205	210	1223
TCC TGG GAA CTG GGC AAT GAG CCC AAC AGT TTc TGG AAG AAA GCT Ser Trp Glu Leu Gly Asn Glu Pro Asn Ser Phe Trp Lys Lys Ala 215	220	225	1268
CAC ATT CTC ATC GAT GGG TTG CAG TTA GGA GAA GAC TTT GTG GAG His Ile Leu Ile Asp Gly Leu Gln Leu Gly Glu Asp Phe Val Glu 230	235	240	1313
TTG CAT AAA CTT cTA CAA AGG TCA GCT TTC CAA AAT GCA AAA CTC Leu His Lys Leu Leu Gln Arg Ser Ala Phe Gln Asn Ala Lys Leu 245	250	255	1358
TAT GGT CCT GAC ATC GGT CAG CCT CGA GGG AAG ACA GTT AAA CTG Tyr Gly Pro Asp Ile Gly Gln Pro Arg Gly Lys Thr Val Lys Leu 260	265	270	1403
CTG AGG AGT TTC CTG AAG GCT GGC GGA GAA GTG ATC GAC TCT CTT Leu Arg Ser Phe Leu Lys Ala Gly Gly Glu Val Ile Asp Ser Leu 275	280	285	1448
ACA TGG CAT CAC TAT TAC TTG AAT GGA CGC ATC GCT ACC AAA GAA Thr Trp His His Tyr Tyr Leu Asn Gly Arg Ile Ala Thr Lys Glu 290	295	300	1493
GAT TTT CTG AGC TCT GAT GCG CTG GAC ACT TTT ATT CTC TCT GTG Asp Phe Leu Ser Ser Asp Ala Leu Asp Thr Phe Ile Leu Ser Val 305	310	315	1538
CAA AAA ATT CTG AAG GTC ACT AAA GAG ATC ACA CCT GGC AAG AAG Gln Lys Ile Leu Lys Val Thr Lys Glu Ile Thr Pro Gly Lys Lys 320	325	330	1583
GTC TGG TTG GGA GAG ACG AGC TCA GCT TAC GGT GGC GGT GCA CCC Val Trp Leu Gly Glu Thr Ser Ser Ala Tyr Gly Gly Ala Pro 335	340	345	1628
TTG CTG TCC AAC ACC TTT GCA GCT GGC TTT ATG TGG CTG GAT AAA Leu Leu Ser Asn Thr Phe Ala Ala Gly Phe Met Trp Leu Asp Lys 350	355	360	1673
TTG GGC CTG TCA GCC CAG ATG GGC ATA GAA GTC GTG ATG AGG CAG Leu Gly Leu Ser Ala Gln Met Gly Ile Glu Val Val Met Arg Gln 365	370	375	1718
GTG TTC TTC GGA GCA GGC AAC TAC CAC TTA GTG GAT GAA AAC TTT Val Phe Phe Gly Ala Gly Asn Tyr His Leu Val Asp Glu Asn Phe 380	385	390	1763
GAG CCT TTA CCT GAT TAC TGG CTC TCT CTT CTG TTC AAG AAA CTG Glu Pro Leu Pro Asp Tyr Trp Leu Ser Leu Leu Phe Lys Lys Leu 395	400	405	1808

GTA GGT CCC AGG GTG TTA CTG TCA AGA GTG AAA GGC CCA GAC AGG	1853	
Val Gly Pro Arg Val Leu Leu Ser Arg Val Lys Gly Pro Asp Arg		
410	415	420
AGC AAA CTC CGA GTG TAT CTC CAC TGC ACT AAC GTC TAT CAC CCA	1898	
Ser Lys Leu Arg Val Tyr Leu His Cys Thr Asn Val Tyr His Pro		
425	430	435
CGA TAT CAG GAA GGA GAT CTA ACT CTG TAT GTC CTG AAC CTC CAT	1943	
Arg Tyr Gln Glu Gly Asp Leu Thr Leu Tyr Val Leu Asn Leu His		
440	445	450
AAT GTC ACC AAG CAC TTG AAG GTA CCG CCT CCG TTG TTC AGG AAA	1988	
Asn Val Thr Lys His Leu Lys Val Pro Pro Leu Phe Arg Lys		
455	460	465
CCA GTG GAT ACG TAC CTT CTG AAG CCT TCG GGG CCG GAT GGA TTA	2033	
Pro Val Asp Thr Tyr Leu Leu Lys Pro Ser Gly Pro Asp Gly Leu		
470	475	480
CTT TCC AAA TCT GTC CAA CTG AAC GGT CAA ATT CTG AAG ATG GTG	2078	
Leu Ser Lys Ser Val Gln Leu Asn Gly Gln Ile Leu Lys Met Val		
485	490	495
GAT GAG CAG ACC CTG CCA GCT TTG ACA GAA AAA CCT CTC CCC GCA	2123	
Asp Glu Gln Thr Leu Pro Ala Leu Thr Glu Lys Pro Leu Pro Ala		
500	505	510
GGA AGT GCA CTA AGC CTG CCT GCC TTT TCC TAT GGT TTT TTT GTC	2168	
Gly Ser Ala Leu Ser Leu Pro Ala Phe Ser Tyr Gly Phe Phe Val		
515	520	525
ATA AGA AAT GCC AAA ATC GCT GCT TGT ATA TGA AAA TAA AAG GCA	2213	
Ile Arg Asn Ala Lys Ile Ala Ala Cys Ile		
530	535	
TAC GGT ACC CCT GAG ACA AAA GCC GAG GGG GGT GTT ATT CAT AAA	2258	
ACA AAA CCC TAG TTT AGG AGG CCA CCT CCT TGC CGA GTT CCA GAG	2303	
CTT CGG GAG GGT GGG GTA CAC TTC AGT ATT ACA TTC AGT GTG GTG	2348	
TTC TCT CTA AGA AGA ATA CTG CAG GTG GTG ACA GTT AAT AGC ACT	2393	
GTG	2396	
(2) INFORMATION FOR SEQ ID NO:46:		
(i) SEQUENCE CHARACTERISTICS:		
(A) LENGTH: 385		
(B) TYPE: nucleic acid		
(C) STRANDEDNESS: double		
(D) TOPOLOGY: linear		
(xi) SEQUENCE DESCRIPTION: SEQ ID NO:46		
CGGCCGCTGC TGCTGCTGTG GCTCTGGGG CGGCTCCGTG CCCTGACCCA	50	
AGGCCACTCCG CGGGGGACCG CGCCGACCAA AGACGTGGTG GACTTGGAGT	100	
TTTACACCAA GAGGCTATTC CAAAGCGTGA GTCCCTCGTT CCTGTCCATC	150	
ACCATCGACG CCAGTCTGGC CACCGACCCCT CGGTTCCCTCA CCTTCCTGAG	200	
CTCTCCACGG CTTCGAGCCC TGTCTAGAGG CTTATCTCCT GCGTACTTGA	250	
GATTGGCGG CACCAAGACT GACTTCCTTA TTTTGATCC CAACAACGAA	300	
CCCACCTCTG AAGAAAGAAG TTACTGGCAA TCTCAAGACA ACAATGATAT	350	

36

TTGCGGGTCT GACCGGGTCT CCGCTGACGT GTTGA

385

(2) INFORMATION FOR SEQ ID NO:47:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 541
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: double
- (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:47

AAATCAGGAC ATATCCTTCA CTTATTGCC TCTTGGTCAT ATTGGAGGCA	50
TTTGTATTCA TTTTTAAATAA CCCTCAAAAT AGTGCATGCA AAGTGCTAAG	100
CGTCATTGTC CACATGGTGC CATTAACTGT CACCACCTGC AGTGGTCTAC	150
TTAGAGAACCA CCGCACTGGA TGTTAACACT GAAGCGCGTG CCCCCGCCCTC	200
CCGAGGCTCT GGATCCAGCG TTGAAGCTTG CCCCCCCCTC CCGAGGCTCT	250
GGATCCAGCA CTGGAGCATG CCCCCCCCTC CCGAGGCTCT GGAGCTTGCT	300
AAGGAGTCGG CTCCCTACCG CTGGGGTTTT GCTTTATTCT TATGAATGAC	350
ACCCCTGACC GCTTTCGTCT CAGGGGTACT GTAATGCCTT TTATTTTCAT	400
ATACAAGCTG CGATTTGGC ATTTCTTATG ACAAAAAACC CATAGGAAAA	450
GGCGGGCAGC CTTAGTGAGC TTCTGCGGG GAGAGGTTTT TCTGTTAGAG	500
CTGGCANGGT CTGCTCATCG ACCATCTTCA GGCCTCGTGC C	541

INTERNATIONAL SEARCH REPORT

International application No.
PCT/US00/03542

A. CLASSIFICATION OF SUBJECT MATTER

IPC(7) :C12N 15/56, 15/63, 1/21, 9/24, 15/11
US CL :536/23.1, 23.2; 435/200, 325, 252.3, 320.1; 424/94.61

According to International Patent Classification (IPC) or to both national classification and IPC

B. FIELDS SEARCHED

Minimum documentation searched (classification system followed by classification symbols)

U.S. : 536/23.1, 23.2; 435/200, 325, 252.3, 320.1; 424/94.61

Documentation searched other than minimum documentation to the extent that such documents are included in the fields searched

Electronic data base consulted during the international search (name of data base and, where practicable, search terms used)

EAST, MEDLINE, BIOSIS, CAPLUS, SCISEARCH, EMBASE, JAPIO, PATOWEP, PATOSWO search terms: heparanase, gene or sequence

C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
X	US 5,362,641 A (FUKS et al.) 08 November 1994, see entire document.	21-25 ----
-		1-20, 26-28
Y		
X, P	US 5,968,822 A (PECKER et al.) 19 October 1999, see entire document.	1-28
X	WO 95/04158 A1 (THE UPJOHN COMPANY) 09 February 1995, see entire document.	21-25 ----
-		1-24, 26-28
Y		
X, P	WO 99/11798 A1 (INSIGHT STRATEGY & MARKETING LTD.) 11 March 1999, see entire document.	1-28

Further documents are listed in the continuation of Box C. See patent family annex.

Special categories of cited documents:	
"A"	document defining the general state of the art which is not considered to be of particular relevance
"T"	later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention
"B"	earlier document published on or after the international filing date
"X"	document of particular relevance; the claimed invention cannot be considered novel or cannot be considered to involve an inventive step when the document is taken alone
"C"	document which may throw doubts on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified)
"Y"	document of particular relevance; the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art
"O"	document referring to an oral disclosure, use, exhibition or other means
"Z"	document member of the same patent family
"P"	document published prior to the international filing date but later than the priority date claimed

Date of the actual completion of the international search	Date of mailing of the international search report
12 JUNE 2000	24 JUL 2000
Name and mailing address of the ISA/US Commissioner of Patents and Trademarks Box PCT Washington, D.C. 20231	Authorized officer <i>Ruthie Lawrence</i> <i>for</i> RICHARD HUTSON
Facsimile No. (703) 305-3230	Telephone No. (703) 308-0196

INTERNATIONAL SEARCH REPORT

International application No.
PCT/US00/03542

Box I Observations where certain claims were found unsearchable (Continuation of item 1 of first sheet)

This international report has not been established in respect of certain claims under Article 17(2)(a) for the following reasons:

1. Claims Nos.:
because they relate to subject matter not required to be searched by this Authority, namely:

2. Claims Nos.:
because they relate to parts of the international application that do not comply with the prescribed requirements to such an extent that no meaningful international search can be carried out, specifically:

3. Claims Nos.:
because they are dependent claims and are not drafted in accordance with the second and third sentences of Rule 6.4(a).

Box II Observations where unity of invention is lacking (Continuation of item 2 of first sheet)

This International Searching Authority found multiple inventions in this international application, as follows:

Please See Extra Sheet.

1. As all required additional search fees were timely paid by the applicant, this international search report covers all searchable claims.
2. As all searchable claims could be searched without effort justifying an additional fee, this Authority did not invite payment of any additional fee.
3. As only some of the required additional search fees were timely paid by the applicant, this international search report covers only those claims for which fees were paid, specifically claims Nos.:

4. No required additional search fees were timely paid by the applicant. Consequently, this international search report is restricted to the invention first mentioned in the claims; it is covered by claims Nos.:

Remark on Protest

The additional search fees were accompanied by the applicant's protest.

No protest accompanied the payment of additional search fees.

INTERNATIONAL SEARCH REPORT

International application No.

PCT/US00/03542

BOX II. OBSERVATIONS WHERE UNITY OF INVENTION WAS LACKING
This ISA found multiple inventions as follows:

This application contains the following inventions or groups of inventions which are not so linked as to form a single inventive concept under PCT Rule 13.1. In order for all inventions to be searched, the appropriate additional search fees must be paid.

Group I, claim(s)1-7, 19, 20 and 28, drawn to a nucleic acid encoding a polypeptide having heparanase activity.

Group II, claim(s) 8-18, drawn to antisense oligonucleotides of a polynucleotide which encodes a polypeptide having heparanase activity.

Group III, claim(s) 21-25, drawn to polypeptide having heparanase activity.

Group IV, claim(s)26, drawn to a method of identifying a chromosome region harboring a heparanase gene.

Group V, claim(s) 27, drawn to a method of eliciting anti-heparanase antibodies in vivo.

The species listed above do not relate to a single inventive concept under PCT Rule 13.1 because, under PCT Rule 13.2, the species lack the same or corresponding special technical features for the following reasons: The listed inventions share a technical relationship of a polypeptide having heparanase activity, but this does not constitute a special technical feature because Fuks et al. (Fuks et al. US Patent No: 5,362,641) teach a polypeptide having heparanase activity.